

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12P 21/00, C12N 15/00 C07K 13/00, C07H 15/12	A1	(11) International Publication Number: WO 91/05057 (43) International Publication Date: 18 April 1991 (18.04.91)
(21) International Application Number: PCT/US90/05520 (22) International Filing Date: 27 September 1990 (27.09.90) (30) Priority data: 413,938 28 September 1989 (28.09.89) US (71) Applicant: THE SALK INSTITUTE BIOTECHNOLOGY/INDUSTRIAL ASSOCIATES, INC. [US/US]; 10280 North Torrey Pines Road, LaJolla, CA 92037 (US). (72) Inventors: BUCKHOLZ, Richard, Gordon ; 1667 Clearwater Place, Encinitas, CA 92024 (US). BRIERLEY, Russell, Arthur ; 6520 Via Barona, Carlsbad, CA 92009 (US). ODIORNE, Mary, Susan ; 2251 Altisma Way #101, Carlsbad, CA 92009 (US). SIEGEL, Robert, Steven ; 2633 Wyandotte, San Diego, CA 92117 (US). WONDRACK, Lillian, Margaret ; 28 Corey Lane, Niantic, CT 06357 (US).		(74) Agents: WATT, Phillip, H. et al.; Fitch, Even, Tabin & Flannery, Room 900, 135 South LaSalle Street, Chicago, IL 60603 (US). (81) Designated States: AT (European patent), BE (European patent), CA, CH (European patent), DE (European patent)*, DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), NO, SE (European patent). Published <i>With international search report.</i>
(54) Title: CD4 PRODUCTION IN PICHIA PASTORIS (57) Abstract <p>The recombinant production of a secreted, soluble form of CD4 protein (a glycoprotein that is expressed on the cell membrane, thymus-derived (T) lymphocytes) is described, employing a <i>P. pastoris</i> expression system. Also disclosed are DNA fragments which are useful for the expression of CD4 in <i>P. pastoris</i>, expression vectors containing such DNA fragments, yeast cells transformed with such DNA fragments or vectors, and the substantially pure human CD4 glycoprotein produced thereby.</p>		

DESIGNATIONS OF "DE"

Until further notice, any designation of "DE" in any international application whose international filing date is prior to October 3, 1990, shall have effect in the territory of the Federal Republic of Germany with the exception of the territory of the former German Democratic Republic.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MC	Monaco
AU	Australia	FI	Finland	MG	Madagascar
BB	Barbados	FR	France	ML	Mali
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Fasso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GR	Greece	NL	Netherlands
BJ	Benin	HU	Hungary	NO	Norway
BR	Brazil	IT	Italy	PL	Poland
CA	Canada	JP	Japan	RO	Romania
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
DE	Germany	LU	Luxembourg	TD	Chad
DK	Denmark			TG	Togo
				US	United States of America

CD4 PRODUCTION IN PICHIA PASTORISField of the Invention

This invention relates to the field of recombinant DNA technology. More particularly, the invention concerns the development of Pichia pastoris yeast strains capable of high-level production and secretion of at least a portion of the human T-cell receptor molecule CD4 (also referred to as T4 protein) containing the site of interaction between CD4 and the human immunodeficiency virus HIV.

10

Background of the Invention

The CD4 protein is a glycoprotein of approximately 60,000 daltons molecular weight that is expressed on the cell membrane of the mature, thymus-derived (T) lymphocytes, and to a lesser extent on cells of the monocyte/macrophage lineage. The CD4 molecule consists of four tandem extracellular domains which contain significant sequence and structural homology with the variable (V) and joining (J) regions of immunoglobulin gene family members, a single membrane-spanning domain, and a carboxy-terminal cytoplasmic segment.

15

20

The molecule was originally described as a marker distinguishing the helper/inducer subset of mature T lymphocytes [Reinherz et al., Cell 19, 821 (1980); Goldstein et al., Immunol. Review 68, 5 (1982)], and is known to be involved in the interaction of these cells with components of the immune system that express class II major histocompatibility complex (MHC) antigen molecules [see, for example, Swain, Immunol. Review 74, 129 (1983); Gay et al., Nature 328, 626 (1987); Doyle et al., Nature 330, 256 (1987)].

25

30

The isolation and nucleotide sequence of a cDNA encoding the CD4 surface glycoprotein was first reported by Maddon et al. (Columbia University), Cell 42, 93

35

(1985) and is disclosed in the PCT Patent Application Publication No. WO 88/01304. However, the published sequence proved to be incorrect at its N-terminus due to a sequencing error in which the AAG codon (nucleotides 151-153 in the cDNA clone pT4B) was reported as AAC. Accordingly, the authors originally predicted an asparagine (asn) at the +3 position. Subsequently, Richard Axel's group at Columbia University resequenced the pT4B cDNA and sequenced three cDNAs from different libraries, as well as genomic clones encoding CD4. They have found that CD4 actually contains lysine (lys) at the asn assignment, and that the residue designated originally as +3 is, in fact, the amino-terminal residue [Littman *et al.* in *Cell* 55, 541 (1988)].

Of immediate interest, is the finding that the human CD4 protein binds the human immunodeficiency virus (HIV), the causative agent of AIDS; and it is believed that the HIV virus gains entry to the cells through interaction with the CD4 "receptor". Amongst the earliest publications concerning the interaction of the CD4 molecule and the HIV virus are, for example: Dalglish *et al.*, *Nature* 312, 763 (1984); Klatzman *et al.*, *Nature* 312, 767 (1984); McDougal *et al.*, *J. Immunol.* 135, 3151 (1985); and McDougal *et al.*, *Science* 231, 382 (1986).

Recent reports have described transfection of CD4⁻ cells with CD4-encoding DNA, and the subsequent newly acquired ability of transformed cells to bind HIV and become infected. Thus, Maddon *et al.*, *Cell* 47, 333 (1986) described the recombinant expression of the CD4 (T4) gene in human lymphoid and epithelial cells, and the ability of previously T4⁻ cells to bind to and become infected with HIV, after they had been transformed with recombinant vectors and thereby became T4⁺ cells. The authors also showed that recombinant human CD4 on mouse cells did not allow for HIV infection, although it bound HIV.

Further publications concerning the recombinant production of soluble CD4 protein, and the ability of the recombinant product to interact with HIV and thus inhibit infection are: Smith et al., Science 238, 1704 (1987);
5 Fisher et al., Nature 331, 76 (1988); Hussey et al.,
Nature 331, 78 (1988); Deen et al., Nature 331, 82
(1988); and Traunecker et al., Nature 331, 84 (1988). A
concise review on the subject of HIV/CD4 interaction has,
for example, been published by Q. J. Sattentau and R. A.
10 Weiss in Cell 52, 631 (1988).

Smith et al., Supra produced soluble, secreted forms of the CD4 antigen molecule by transfection of mammalian (CHO) cells with vectors encoding truncated versions of CD4, in which the transmembrane and
15 cytoplasmic domains were replaced with a short linker sequence containing an in-frame stop codon. The authors worked under the assumption that the deduced amino acid sequence of CD4 as originally published by Maddon et al., was correct.

20 Fisher et al., Supra constructed three truncated CD4 genes that lacked the transmembrane and cytoplasmic domains, and produced recombinant soluble CD4 protein in dihydrofolate reductase (DHFR)-mutant CHO cells. The authors discovered the discrepancy in the CD4
25 amino acid sequence when they sequenced their own cDNA, but attributed it to a possible allelic polymorphism and chemically changed the AAG codon to an AAC codon to obtain a CD4 protein sequence "identical to that previously reported" (page 331).

30 To produce a secreted form of CD4, Hussey et al., Supra report the expression of truncated CD4 gene in Spodoptera frugiperda (SF) cells, using a baculovirus (AcNPV) expression system. Milligram quantities of a hydrophilic extracellular segment of CD4 were generated.

35 Deen et al., Supra described an expression system in which a recombinant, soluble form of CD4 was secreted into tissue culture supernatants. Supernatants

from clones were monitored for the expression of soluble CD4, among others, by Western blot analysis using a rabbit anti-CD4 polyclonal antibody "developed against a denatured CD4 protein produced in bacteria" (page 82).

5 Trauneker et al., Supra report the production and secretion of two soluble chimeric CD4 proteins from myeloma cells. The secreted proteins retained "at least some of their original conformation" (page 84).

10 The specific sequences of CD4 and the HIV virus that are required for interaction have also been identified.

15 The component of HIV that mediates binding of the virus to CD4 is the surface glycoprotein, gp120 [Lasky et al., Cell 50, 975 (1987)]. To date, antibodies raised against gp120 have been ineffective in blocking viral infection either in vitro or in vivo. The inability to block is probably related to the heterogeneity seen among gp120 protein sequences from different viral isolates. Antibodies raised against gp120 from one HIV isolate will not necessarily recognize gp120 from a different isolate. Also, the CD4-binding region of gp120 is not accessible to antibody molecules and thus may be capable of binding CD4 even if antibody does bind gp120.

20 Studies using monoclonal antibodies to CD4 have identified the first variable region, V₁, comprising the N-terminal 106 amino acids of mature CD4, as the site of interaction with gp120 [Berger et al., PNAS 85, 2357 (1988)]. Further analyses of binding, using mutant CD4 proteins, truncated derivatives of CD4, HIV, and purified gp120 have narrowed this assignment to amino acid residues 40-48 within V₁ [Peterson and Seed, Cell 54, 65 (1988)]. However, other conserved structures within V₁ are probably essential to achieve the highest affinity binding of HIV to CD4.

35 The affinity of the HIV virus for CD4 makes this molecule a rational target for development of an effective AIDS therapy or prevention. It might be

possible to block the entry of HIV into CD4-expressing T-cells through the use of anti-CD4 antibodies or the presence of excess soluble CD4 molecules. In the case of antibodies, the CD4 receptor present on the T-cell surface may be unable to bind the HIV gp120 "ligand" if the CD4 is first bound by antibody. Alternatively, if an excess of soluble CD4 molecules is present in a sample comprised of CD4-expressing T-cells and HIV, then a large proportion of virus might bind to the soluble CD4, be inhibited from binding the cell-associated receptor, and viral infection might be lessened or prohibited.

Chao *et al.*, *J. Biol. Chem.* 264, 5812 (1989) expressed a gene encoding a 113-amino acid, NH₂-terminal fragment of CD4 (rST4.113) in *E. coli* under the control of the *E. coli* tryptophan operon promoter. An insoluble product that is found in inclusion bodies, is obtained at 5 to 10% of total protein, the purification of which provides the recombinant peptide at less than 20% of the starting material. The product, unlike the naturally occurring CD4 contains an unblocked N-terminal methionine group.

In view of the promising therapeutic results, there is a great need for a recombinant expression system that is suitable for the efficient, large-scale production of a soluble, authentic form of the CD4 protein, that, after purification, is suitable for use in possible AIDS therapies and preventive measures.

The *Pichia pastoris* yeast expression system, developed in part by scientists at SIBIA, the assignee of the present patent application, has proved to be instrumental in the production of several heterologous proteins. This system is based on methanol-regulated promoters and high cell density fermentation. Because *Pichia pastoris* is a methylotrophic yeast, it has metabolic pathways that respond to and regulate methanol utilization. A key enzyme in the methanol utilization pathway is alcohol oxidase, a protein encoded by two

genes, AOX1 and AOX2. When Pichia cells are grown in the presence of methanol, the AOX1 and AOX2 genes are transcribed and a large amount of alcohol oxidase protein is produced. The high level of AOX gene expression is
5 mainly due to the methanol-responsive AOX1 gene promoter which is activated in the presence of methanol. This promoter is highly expressed and tightly regulated (see e.g. the European Patent Application No. 85113737.2, published June 4, 1986, under No. 183 071). After
10 identification and isolation of the AOX1 regulatory elements, a methanol-responsive gene expression system has been developed in Pichia that places heterologous genes under the regulation of the AOX1 promoter [Cregg et al., Bio/Technology 5, 479 (1987)]. Another key
15 feature of the P. pastoris expression system is the stable integration of expression cassettes into the P. pastoris genome, thus significantly decreasing the chance of vector loss.

Although P. pastoris has been used successfully
20 for the production of various heterologous proteins, e.g., hepatitis B surface antigen [Cregg et al., Supra], bovine lysozyme [Digan et al., Developments in Industrial Microbiology 29, 59 (1988); Digan et al., Bio/Technology 7, 160 (1989), and Saccharomyces cerevisiae invertase
25 [Tschopp et al., Bio/Technology 5, 1305 (1987)], endeavors to produce other heterologous gene products in Pichia, especially by secretion, have given mixed results and, in some cases, have been unsuccessful. At our present level of understanding of the P. pastoris
30 expression system, it is unpredictable whether a given gene can be expressed to an appreciable level in this yeast, whether the expression yields a product that is stable under ordinary fermentation conditions and subsequent processing, or whether Pichia will tolerate
35 the presence of the recombinant gene product in its cells. Further, it is especially difficult to foresee if a particular protein will be secreted by P. pastoris, and

if it is, at what efficiency. Even for S. cerevisiae, which has been considerably more extensively studied than P. pastoris, the mechanism of protein secretion is not well defined and understood.

5

Summary of the Invention

The present invention relates to the production of a secreted soluble form of CD4 protein, containing the site of interaction between CD4 and the human immunodeficiency virus (HIV) in Pichia pastoris (P. pastoris).

10

In one aspect, the present invention relates to a P. pastoris yeast cell containing in its genome at least one copy of a DNA sequence operably encoding in P. pastoris at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and HIV, in operational association with a DNA sequence encoding a signal sequence which functions to direct secretion of the encoded glycoprotein in P. pastoris, both under the regulation of a promoter region of a P. pastoris gene. The signal sequence of the S. cerevisiae alpha-mating factor (AMF) gene (AMF pre-pro sequence) is a preferred signal sequence.

15

20

In another aspect, the present invention concerns a DNA fragment which may be contained within, or may itself be, a circular plasmid, and which comprises at least one copy of an expression cassette comprising in the direction of transcription, a promoter region of a first P. pastoris gene, a DNA sequence encoding in P. pastoris at least a portion of human CD4 glycoprotein containing the site of interaction between CD4 and the HIV virus, preceded by a DNA sequence encoding a signal sequence directing the secretion of said glycoprotein or a portion thereof in P. pastoris, and a transcription terminator of a second P. pastoris gene, said first and second P. pastoris genes being identical or different, and the segments of said expression cassette being in

30

35

operational association. The DNA sequence preceeding the CD4 glycoprotein gene preferably is a DNA sequence encoding the S. cerevisiae AMF pre-pro sequence followed by a DNA sequence encoding AMF processing site lys-arg.

5

Expression vectors containing such DNA sequences are also within the scope of the invention.

10

In a further aspect, the invention relates to a process for producing and secreting at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the virus HIV, into the culture medium. According to this process, P. pastoris transformants containing in their genome at least one copy of a DNA sequence operably encoding in P. pastoris at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the virus HIV, in operational association with a DNA sequence encoding a signal sequence which functions to direct secretion of the encoded CD4 or CD4 portion in P. pastoris (the S. cerevisiae AMF pre-pro sequence being preferred), both under the regulation of a promoter region of a P. pastoris gene, are grown under conditions allowing the expression of the DNA sequences in P. pastoris and secretion of the CD4 glycoprotein into the culture medium in a substantially pure form devoid of degradation products.

15

20

25

Brief Description of Drawings

30

Figure 1 shows the nucleotide sequence and amino acid sequence of the S. cerevisiae alpha-mating factor (AMF) pre-pro gene segment.

Figure 2 shows the nucleotide sequence and the deduced amino acid sequence of a 482 bp DNA fragment encoding amino acids 1 - 106 of mature CD4 along with its leader sequence.

Figure 3 illustrates the construction of the Pichia pastoris expression vector, pSCD103 for the production of human CD4-V₁.

Figure 4 shows the nucleotide and amino acid sequence of the EcoRI insert of pSCD103.

Figure 5 is a restriction map of plasmid pAO815

Figure 6 shows the cell wet weight over time for fermentation Runs 568, 570, and 571.

Figure 7 shows the time course of fermentation Run 593 of the two-copy Mut⁺ strain, G+SCD103S16.

- A. Cell density (grams of wet weight/liter), plotted against time of fermentation.
- B. Recombinant human CD4-V₁ production (mg/liter of cell-free fermentor broth) for the fermentation presented in Figure 7A is plotted against time. The expression level was determined by the quantitative Western blot assay.

Figure 8 is a silver-stained gel, using reducing conditions of V₁ standard and Pichia pastoris fermentor broth. Lanes 1-4 (numbered consecutively from left to right) contain 100, 200, 300 or 400 ng V₁ standard, respectively. Lane 6 is 7.5 μ l of G+SCD103S16 fermentor broth; lanes 7 and 8 are 5 μ l of G-PAO815 fermentor broth containing 100 ng or 200 ng V₁ standard, respectively. Lane 10 is 3.75 μ l of G+SCD103S16 broth with 3.75 μ l of G-PAO815 broth; lanes 11 and 12 are 3.75 μ l and 7.5 μ l of G-PAO815 broth, alone. Lane 13 is pre-stained molecular size standards obtained from Diversified Biotech, Newton Centre, MA. They are Low Range Standards #SDS-100P and are: trypsin inhibitor, 20,400; myoglobin, 16,949; CNBr cleavage fragments of myoglobin, Fragment IV, 14,404; Fragment III, 8,159; Fragment II, 6,214; Fragment I 2,152. These standards are included to denote relative positions of sample bands on the gel and have not been used to estimate molecular mass. Lanes 5 and 9 contain broth from a fermentation

run of strain G+SCD103S16 using slightly different conditions than those reported in the Examples.

Figure 9 shows the results of N-terminal sequence analysis of rCD4-V₁ produced in Pichia pastoris compared with the published N-terminal sequence for mature human CD4. Pichia rCD4-V₁ was isolated from a gel similar to the one in Figure 8, and sequenced as described in Example 4b.

Figure 10 shows the result of stability test performed on rCD4-V₁ produced in Pichia pastoris. One hundred microliter samples of G+SCD103S16 broth were treated under the following conditions, before 10 μ l was separated on SDS-PAGE and subjected to immunoblotting with polyclonal sCD4 antibody. All broth samples were frozen immediately upon removal from the fermentor. The sample in lane 1 was thawed just prior to SDS-PAGE. Lanes 2-4 contain samples at pH 2.5, lanes 5-7 contain samples that had been adjusted to pH 5.0 upon thawing. Samples in: lanes 2 and 5 were thawed and immediately refrozen, lanes 3 and 6 were thawed and incubated at 4° for 20 hours, lanes 4 and 7 were thawed and incubated at 30°C for 20 hours. E. coli V₁ (100 ng) was included as a size standard.

Figure 11 shows the nucleotide sequence of the human CD4 cDNA and the translated sequence of the CD4 protein.

Detailed Description of the Invention

1. Definitions

An expression system suitable for the production and secretion of at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the human immunodeficiency virus HIV is provided. Preferably, this portion is the V₁ region of CD4.

It will be understood that there is some uncertainty in the literature as to the definition of the "first variable region" (V_1 region) of CD4, often referred to as "CD4- V_1 ". Studies using recombinant HIV gp120 demonstrate that the determinants for high affinity binding lie solely within the first 106 amino acids of CD4. However, the "recombinant V_1 " produced in *E. coli* contains the first 113 N-terminal amino acids of mature human CD4. The terms "first variable region", " V_1 " or " V_1 region" and synonymous expressions, alone or in combination with other terms, are used throughout the specification and claims to refer to a DNA sequence including at least the first 106 N-terminal amino acids of mature human CD4 (as shown in Figure 11). However, polypeptides deficient in one or more amino acids in the amino acid sequence reported in the literature, or polypeptides containing additional amino acids, or polypeptides in which one or more amino acids in the amino acid sequence of the V_1 region of CD4 are replaced by other amino acids, are within the scope of the definition used herein, provided that they exhibit the functional activity of V_1 , in particular preserve its HIV-binding properties. The definition used in connection with the present invention is intended to embrace all the allelic variations of V_1 . Moreover, as noted *Supra*, derivatives obtained by simple modification of the amino acid sequence of the naturally occurring product, e.g. by way of site-directed mutagenesis or other standard procedures are included.

The term "at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the human immunodeficiency virus HIV" and synonymous expressions, as used herein, refer to the full-length mature human CD4 glycoprotein molecule or any portion thereof capable of binding the HIV virus. Just as in the case of the V_1 region of CD4, polypeptides deficient in one or more amino acids in the correct amino

acid sequence reported in the literature for mature human naturally occurring CD4 or its respective regions, or polypeptides containing additional amino acids, or polypeptides in which one or more amino acids in the amino acid sequence of CD4 or its respective regions are replaced by other amino acids, are within the scope of the definition used herein, provided that they exhibit the functional activity of CD4, in particular preserve its HIV-binding properties. The definition used in connection with the present invention is intended to embrace all the allelic variations of CD4. Moreover, as noted Supra, derivatives obtained by simple modification of the amino acid sequence of the naturally occurring protein, e.g. by way of site-directed mutagenesis or other standard procedures are included.

The term "amino acid sequence operably encoding in Pichia pastoris at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the human immunodeficiency virus HIV" and grammatical variations thereof, as used herein, refers to DNA sequences encoding in Pichia pastoris "at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the human immunodeficiency virus HIV" such as the "first variable (V_1) region", as hereinabove defined. This sequence contains but is not restricted to, the DNA sequence encoding residues 16 through 84 of the mature CD4 protein, which are contained within the first disulfide-bonded, covalently closed peptide loop of CD4. Such sequences may be obtained by chemical synthesis or by transcription of a messenger RNA (mRNA) corresponding to CD4 or a portion thereof to a complementary DNA (cDNA) and converting the latter into a double stranded cDNA. Additionally, the CD4 sequences may be obtained through the use of polymerase chain reaction (PCR) on genomic DNA encoding at least the V_1 region. The mRNA can be isolated for example, from $T4^+$ transformed fibroblasts as described by Maddon et al.,

Supra (1985). Chemical synthesis of a gene for human CD4 or a portion thereof is, for example, disclosed by Jameson et al., Science 240, 1335 (1988); Litsen et al., Science 241, 712 (1988). The requisite DNA sequence can also be removed, for example, by restriction enzyme digest of known vectors harboring the desired gene. Examples of such vectors and the means for their preparation can be taken from the following publications: Smith et al., Supra; Fisher et al., Supra; Hussey et al., Supra, Deen et al., Supra, Traunecker et al., Supra, etc. According to Example 1 of the present application, a 482 bp DNA fragment encoding the V₁ portion of human CD4 was excised from a 2.2 kb linear BglII-NheI DNA fragment by digestion with EcoRI and XbaI. However, the CD4-V₁ encoding DNA fragment can be removed from other known DNA fragments as well. For example, a 682 bp EcoRI-NheI fragment from pT4B is disclosed in Maddon et al., Cell 42, 93 (1985). The V₁-encoding sequence can be readily obtained from this fragment by digestion with EcoRI and XbaI.

The amino acids, which occur in the various amino acid sequences referred to in the specification have their usual, three- and one-letter abbreviations, routinely used in the art, i.e.:

25	<u>Amino Acid</u>	<u>Abbreviation</u>	
	L-Alanine	Ala	A
	L-Arginine	Arg	R
	L-Asparagine	Asn	N
	L-Aspartic acid	Asp	D
30	L-Cysteine	Cys	C
	L-Glutamine	Gln	Q
	L-Glutamic Acid	Glu	E
	L-Glycine	Gly	G
	L-Histidine	His	H
35	L-Isoleucine	Ile	I
	L-Leucine	Leu	L
	L-Lysine	Lys	K

	L-Methionine	Met	M
	L-Phenylalanine	Phe	F
	L-Proline	Pro	P
	L-Serine	Ser	S
<u>5</u>	L-Threonine	Thr	T
	L-Tryptophan	Trp	W
	L-Tyrosine	Tyr	Y
	L-Valine	Val	V

10 The promoter region employed to drive the expression of a gene encoding at least a portion of CD4, preferably CD4-V₁, is derived from a methanol-regulated alcohol oxidase gene of P. pastoris. P. pastoris is known to contain two functional alcohol oxidase genes:

15 alcohol oxidase I (AOX1) and alcohol oxidase II (AOX2) genes. The coding portions of the two AOX genes are closely homologous at the DNA and predicted amino acid sequence levels and share common restriction sites. The proteins expressed from the two genes have similar

20 enzymatic properties but the promoter of the AOX1 gene is more efficient and highly expressed, therefore, its use is preferred for heterologous expression. The AOX1 gene, including its promoter, has been isolated and thoroughly characterized [Ellis et al., Mol. Cell. Biol. 5, 1111

25 (1985)].

The expression cassette used for transforming P. pastoris cells contains, in addition to the P. pastoris promoter and the CD4 (CD4-V₁) encoding DNA sequence, a DNA sequence encoding a signal sequence

30 directing the secretion of the CD4 glycoprotein or a portion thereof in P. pastoris, preferably a DNA encoding the in-reading frame S. cerevisiae AMF pre-pro sequence, and a DNA sequence encoding AMF processing site, lys-arg (also referred to as lys-arg encoding sequence) and a

35 P. pastoris transcription terminator. Although the S. cerevisiae AMF pre-pro sequence is preferred, other signal sequences suitable for directing foreign protein

secretion in P. pastoris may also be used. Such sequences are, for example, the Saccharomyces cerevisiae invertase signal sequence.

5 The S. cerevisiae alpha-mating factor is a 13-residue peptide, secreted by cells of the "alpha" mating type, that acts on cells of the opposite "a" mating type to promote efficient conjugation between the two cell types and thereby formation of "a-alpha" diploid cells [Thorner et al., The Molecular Biology the Yeast
10 Saccharomyces, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 143 (1981)]. The AMF pre-pro sequence is a leader sequence contained in the AMF precursor molecule, which, together with the lys-arg encoding sequence is necessary for proteolytic processing and secretion (see
15 e.g. Brake et al., Supra). The AMF pre-pro sequence, including the lys-arg encoding sequence is a 255 bp fragment which is illustrated in Figure 1.

The P. pastoris transcription terminator used in accordance with the present invention has a subsegment
20 which encodes a polyadenylation signal and polyadenylation site in the transcript and/or a subsegment which provides a transcription termination signal for transcription from the promoter used in the expression cassette according to the invention (the term
25 "expression cassette" as used herein and throughout the specification and claims refers to a DNA sequence which includes sequences functional for expression and the secretion processes). The entire transcription terminator is taken from a P. pastoris protein-encoding
30 gene, which may be the same or different from the P. pastoris gene which is the source of the P. pastoris promoter used according to the invention.

The DNA fragments according to the invention further comprise a selectable marker gene. For this
35 purpose, any selectable marker gene functional in P. pastoris may be employed, i.e., any gene which confers a phenotype upon P. pastoris cells thereby allowing them

to be identified and selectively grown from among a vast majority of untransformed cells. Suitable selectable marker genes include, for example, selectable marker systems composed of an auxotrophic mutant P. pastoris host strain and a wild type biosynthetic gene which complements the host's defect. For transformation of his4⁻ P. pastoris strains, for example, the S. cerevisiae or P. pastoris HIS4 gene, or for transformation of arg4⁻ mutants the S. cerevisiae ARG4 gene or the P. pastoris ARG4 gene, may be employed.

The term "expression vector" includes vectors capable of expressing DNA sequences contained therein, where such sequences are in operational association with other sequences capable of effecting their expression, i.e. promoter sequences. In general, expression vectors usually used in recombinant DNA technology are often in the form of "plasmids", i.e. circular, double-stranded DNA loops which, in their vector form, are not bound to the chromosome. In the present specification the terms "vector" and "plasmid" are used interchangeably. However, the invention is intended to include other forms of expression vectors as well, which function equivalently.

In the DNA fragment according to the invention the segments of the expression cassette are "in operational association". The DNA sequence encoding CD4 or any portion thereof as hereinabove defined, is positioned and oriented functionally with respect to the promoter, the DNA encoding a signal sequence, preferably the S. cerevisiae AMF pre-pro sequence, and the DNA sequence encoding AMF processing-site, lys-arg and the transcription terminator, so that the polypeptide encoding segment is transcribed, under regulation of the promoter region, into a transcript capable of providing, upon translation the desired polypeptide in P. pastoris. Because of the presence of the signal sequence, e.g. the AMF pre-pro sequence, the expressed product, CD4 or a

portion thereof, as hereinabove defined, is found as a secreted entity in the culture medium, properly processed away from the AMF pre-pro sequence. Appropriate reading frame positioning and orientation of the various segments of the expression cassette are within the knowledge of persons of ordinary skill in the art; further details are given in the Examples.

The DNA fragment provided by the present invention may include sequences allowing for its replication and selection in bacteria, especially *E. coli*. In this way, large quantities of the DNA fragment can be produced by replication in bacteria.

The term "culture" means a propagation of cells in a medium conducive to their growth, and all subcultures thereof. The term "subculture" refers to a culture of cells grown from cells of another culture (source culture), or any subculture of the source culture, regardless of the number of subculturings which have been performed between the subculture of interest and the source culture.

The following abbreviations are used throughout the Examples with the following meanings:

DTT	dithiothreitol
SDS	sodium dodecyl sulfate
PBS	phosphate buffered saline
Mut ⁺	methanol utilization competent
Mut ⁻	methanol utilization defective
Tris-HCl:	1.5M at pH 8.8 and 0.5M at pH6.8. Both are stored at 4°C.

The buffers and solutions, the composition of which is not specified in the Examples, are as follows:

WBB: 1x PBS, 0.05% Tween-20, 0.02% NaN₃, 0.25% gelatin.

Laemmli buffer: see Nature 227, 680 (1970)

2. General Methods

Methods of transforming Pichia pastoris as well as methods applicable for culturing P. pastoris cells containing in their genome a gene for a heterologous protein are known generally in the art.

According to the invention, the expression cassettes are generally transformed into the P. pastoris cells by the whole-cell lithium chloride yeast transformation system [Ito et al., Agric. Biol. Chem. 48, 341 (1984)], with minor modification necessary for adaptation to P. pastoris. Alternatively, the spheroplast technique, described by Cregg et al., Mol. Cell. Biol. 5, 3376 (1985) can also be used for the transformation of P. pastoris cells. The whole-cell lithium chloride method is more convenient in that it does not require the generation and maintenance of spheroplasts.

Positive transformants are characterized by Southern blot analysis [E. M. Southern, J. Mol. Biol. 98, 503 (1975); Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA (1982)] for the site of DNA integration, and Northern blots [Alwine et al., Proc. Natl. Acad. Sci. USA 74, 5350 (1977); Maniatis, Op. Cit.] for methanol-responsive heterologous gene expression. Total RNA for Northern blot analysis was prepared essentially as described by Zitomer et al., J. Biol. Chem. 251, 6320 (1976).

Nick translation can be performed according to Meinkoth et al., Methods in Enzymology 152, 91 (1987).

Transformed strains, which are of the desired phenotype and genotype are grown in fermentors. For the large-scale production of recombinant DNA-based products in P. pastoris a three-stage, high cell-density, batch fermentation system is normally employed. In the first, or growth stage expression hosts are cultured in defined minimal medium with excess glycerol as carbon source. On

this carbon source heterologous gene expression is completely repressed, which allows the generation of cell mass in the absence of heterologous protein expression. Next, a period of glycerol limitation growth is allowed to further increase cell density. Subsequent to the glycerol limited growth, methanol is added, initiating the expression of the desired heterologous protein. This third stage is the so-called production stage. The fermentation of CD4-V₁ essentially followed the three-stage protocol described in Digan *et al.*, Bio/Technology 7, 160 (1989). However, as shown in the Examples and in the description of preferred embodiments, in order to obtain a stable product, the pH had to be maintained at a lower level than usual for Pichia pastoris fermentations.

15

3. Description of Preferred Embodiments

According to a preferred embodiment of the present invention, the V₁ region of the human CD4 molecule is produced in Pichia pastoris. This V₁ region contains a single disulfide bond between two cysteine residues, which are located near the N- and C-termini, respectively.

20

The heterologous protein expression system used for CD4-V₁ production preferably utilizes the promoter derived from the methanol-regulated AOX1 gene of P. pastoris, which is very efficiently expressed and tightly regulated. This gene is the source of the transcription terminator as well. The expression cassette preferably comprises, in operational association, a P. pastoris AOX1 promoter, DNA encoding the S. cerevisiae AMF pre-pro sequence, a DNA sequence encoding AMF processing site, lys-arg, a DNA sequence encoding the CD4-V₁ molecule, and a transcription terminator derived from the P. pastoris AOX1 gene.

30

35

The host cells to be transformed with a linear vector comprising the expression cassette are P. pastoris cells having at least one mutation that can be

complemented with a marker gene present on a transforming DNA fragment. Preferably *his4⁻* (GS115) auxotrophic mutant *P. pastoris* strains are employed.

5 The expression cassette is inserted into a plasmid containing a marker gene complementing the host's defect. pBR322-based plasmids, e.g. pA0815 are preferred. Plasmid pA0815 comprising the CD4-V₁ expression/secretion cassette is called pSCD103. The construction of this plasmid is disclosed in
10 Example 1.

To develop expression strains, the expression cassette is preferably integrated into the host genome by means of the homologous sequences present on the transforming DNA. The expression cassette or entire
15 vector is integrated into the host genome by a one-step gene replacement or addition technique. This approach avoids the problems of plasmid instability. As a result of gene replacement *Mut⁻* strains are obtained. *Mut* refers to the methanol-utilization phenotype. In *Mut⁻* strains,
20 the AOX1 gene is replaced with the expression cassette, thus decreasing the transformant's ability to utilize methanol. A slow growth rate on methanol is maintained by expression of the AOX2 gene product. The transformants in which the expression cassette has
25 integrated into the AOX1 locus by site-directed recombination can be identified by first screening for the presence of the complementing gene. This is preferably accomplished by growing the cells in a media lacking the complementing gene product and identifying
30 those cells which are able to grow by nature of expression of the complementing gene. Next, the selected cells are screened for their *Mut* phenotype by growing them in the presence of methanol and monitoring their growth rate.

35 To develop *Mut⁺* strains, the expression cassette preferably is integrated into the host genome by transformation of the GS115 host with SacI linearized

plasmid pSCD103 comprised of the V₁ expression cassette. The integration is by addition at a locus or loci having homology with one or more sequences present on the transformation vector.

5 Positive transformants are characterized by Southern analysis for the site of DNA integration, by Northern analysis for methanol-responsive CD4-V₁ gene expression, and by immunoblot product analysis for the presence of secreted CD4-V₁ in the growth media.

10 P. pastoris strains which have integrated one or multiple copies of plasmid at a desired site are identified by Southern blot analysis. Strains which demonstrate enhanced expression of the heterologous gene may be identified by Northern analysis, and enhanced secretion

15 of the recombinant protein by product analysis.

For Mut⁻ strains the CD4-V₁ production levels were found to be somewhat lower than for Mut⁺ strains, but the difference was not very significant. Mut⁺ P. pastoris strains integrating multiple copies of the expression

20 vector (or of the AMF-V₁ expression cassette) used for transformation, at the AOX1 locus are preferred, since an increase in copy number often increases productivity.

P. pastoris transformants which are identified to have the desired genotype and phenotype are grown in

25 fermentors. Typically a three-step production process is used. Initially, cells are grown on a repressing carbon source, preferably excess glycerol. In this stage the cell mass is generated in absence of expression. Next, a period of glycerol limitation growth is allowed, and then

30 a limiting methanol feed is initiated, resulting in the expression of the V₁ gene driven by the AOX1 promoter.

It has been found that in the usual pH-range used for heterologous protein production in P. pastoris (about pH 5.0) the V₁ product suffers a substantial

35 proteolytic degradation. To avoid or, at least, reduce product degradation, fermentation is preferably performed at pHs below about 3.5, preferably between about pH 2.5

and 3.5, more preferably between about pH 2.5 and 3.0, for example at about pH 2.6. The pH can be adjusted to the desired value by methods known in the art, preferably before the induction of V₁ production.

5 The level of CD4-V₁ secreted into the media can, for example, be determined by quantitative Western blot analysis of the media in parallel with a standard (e.g. an E. coli produced V₁ standard), using reducing or non-reducing conditions.

10 The invention is further illustrated by the following non-limiting examples.

4. Examples

Example 1

15 Vector construction

I. Construction of the expression vector pSCD103

20 The expression vector construction disclosed in the present application was performed using standard procedures, as described, for example in Maniatis et al., Supra, and Davis et al., Basic Methods in Molecular Biology, Elsevier Science Publishing, Inc., New York (1986).

25 A 2.2 kb linear BglIII-NheI DNA fragment containing a segment encoding the V₁ portion of human CD4 accompanied by flanking DNA from E. coli, was obtained from Smith Kline & French Laboratories (U.S.A.) The V₁-encoding sequence was excised from this fragment by digestion with EcoRI and XbaI, and isolating the 482 bp fragment (the sequence encoding amino acids 1-106 of

30 mature CD4 along with its leader sequence) on a 1.3% agarose gel (Figure 2). Fifty nanograms of the 482 bp fragment were ligated to 100 ng of the plasmid pIBI25, previously cut with EcoRI and XbaI. Plasmid pIBI25 was purchased from IBI, New Haven, CT, and contains an f1

35 origin of replication and the T7 promoter. E. coli MC1061 cells [M.J. Casadaban and S.N. Cohen, J. Mol. Biol. 38, 179 (1980)] were transformed with ligation

products and amp^R colonies were selected. Correct plasmid demonstrated a 477bp band upon digestion with EcoRI and XbaI and was called pSCD4.

5 The AMF pre-pro sequence was isolated from M13mp19 α MF pre-pro by digesting with EcoRI and BamHI and isolating the about 267 bp fragment on a 1.3% agarose gel. To prepare plasmid M13mp19 α MF, 15 μ g of plasmid pAO208 (the construction of which is described hereinafter) were digested with HindIII, filled in with
10 Klenow-fragment DNA polymerase, and digested with EcoRI. The digestion was run on a 1.7% agarose gel and the 267 bp fragment comprised of the AMF pre-pro sequence was isolated. The hEGF (human epidermal growth factor) gene and the AMF pre-pro sequence in the same translational
15 direction were inserted into M13mp19, (New England Biolabs), by the following procedure:

10 μ g of M13mp19 were digested with SmaI and EcoRI and the large, about 7240 bp plasmid fragment was isolated on a 0.8% agarose gel. The plasmid fragment and
20 the 267 bp AMF fragment were ligated together by T4 DNA ligase. The ligation mixture was transformed into JM103 cells and DNA from the plaque was characterized. The correct plasmid was called M13mp19 α MF.

25 Twenty five nanograms of the EcoRI-BamHI fragment of M13mp19 α MF pre-pro were ligated to 100 ng of pIBI25 previously cut with EcoRI and BamHI, and the ligation products were transformed into MC1061 cells. Amp^R colonies were selected and the correct plasmid was identified by digestion with EcoRI and BamHI. The
30 correct plasmid demonstrated a 260 bp band, and was called pAMF101 (Figure 3).

pSCD4 was digested with EcoRI, made blunt-ended by treatment with Klenow fragment of E. coli DNA
35 polymerase I, and then digested with XbaI. The 477 bp V₁-encoding fragment was isolated on a 1.2% agarose gel. pAMF101 was likewise digested with BamHI, treated with Klenow fragment of E. coli DNA polymerase I, digested

with XbaI, and then dephosphorylated. Fifty nanograms of the 477 bp V_1 -encoding fragment were ligated to 100 ng of the linearized vector, and the ligation was transformed into E. coli CJ236 cells (BioRad, Richmond, CA; Muta-gene mutagenesis kit, # 170-3571). Amp^R colonies were selected. The correct plasmid exhibited a 740 bp band upon digestion with EcoRI and XbaI and was called pSCD101 (Figure 3).

Mutagenesis was performed to fuse the AMF pre-pro sequences directly to the V_1 coding region; the STE2 processing sites (glu-ala-glu-ala) of the AMF pre-pro sequence and the native CD4 leader sequence were eliminated by the oligonucleotide-directed mutagenesis. Single-stranded pSCD101 template was prepared following the procedure of Russel *et al.*, Gene 45, 333 (1986), using the helper phage R408. The mutagenizing and screening oligonucleotide was of the following sequence:

5' GGG TAT CTT TGG ATA AAA GAA AGA AAG TGG
TGC TGG GCA A 3'

Mutagenesis reaction products were transformed into MC1061 cells; colonies transformed with the mutagenized plasmid were first identified by hybridization with the screening oligonucleotide, and then the correct mutagenesis was confirmed by sequencing. The correctly mutagenized plasmid was called pSCD102.

An EcoRI linker was added to the 3' end of the AMF pre-pro- V_1 insert by digesting pSCD102 with XbaI, blunt-ending with E. coli DNA polymerase I Klenow fragment, and ligating 100 ng of the vector to 15 ng of EcoRI linkers having the sequence:

5' GGAATTCC 3'

The ligation products were digested with EcoRI and the 560 bp AMF pre-pro- V_1 fragment was isolated on a 1.2% agarose gel. Twenty nanograms of the 560 bp fragment were then ligated to 100 ng of EcoRI-digested and phosphatase-treated pA0815 (the construction of which is described hereinbelow). The ligation products were

transformed into MC1061 cells and the amp^R colonies were selected. The correct plasmid demonstrated a 1675 bp band upon digestion with PstI, and was called pSCD103 (Figure 3). The nucleotide and amino acid sequence of the EcoRI insert of pSCD103 is shown in Figure 4.

II. Construction of plasmid pA0208:

The AOX1 transcription terminator was isolated from 20 µg of pPG2.0 [pPG2.0 = BamHI-HindIII fragment of pG4.0 (NRRL 15868) + pBR322] by StuI digestion followed by the addition of 0.2 µg SalI linkers (GGTCCGACC). The plasmid was subsequently digested with HindIII and the 350 bp fragment isolated from a 10% acrylamide gel and subcloned into pUC18 (Boehringer Mannheim) digested with HindIII and SalI. The ligation mix was transformed into JM103 cells (that are widely available) and amp^R colonies were selected. The correct construction was verified by HindIII and SalI digestion, which yielded a 350 bp fragment, and was called pA0201.

5 µg of pA0201 was digested with HindIII, filled in using E. coli DNA Polymerase I Klenow fragment, and 0.1 µg of BglII linkers (GAGATCTC) were added. After digestion of the excess BglII linkers, the plasmid was reclosed and transformed into MC1061 cells. Amp^R cells were selected, DNA was prepared, and the correct plasmid was verified by BglII, SalI double digests, yielding a 350 bp fragment, and by a HindIII digest to show loss of HindIII site. This plasmid was called pA0202.

The alpha factor-GRF fusion was isolated as a 360 bp BamHI-PstI partial digest from pYSV201. Plasmid pYSV201 is the EcoRI-BamHI fragment of GRF-E-3 inserted into M13mp18 (New England Biolabs). Plasmid GRF-E-3 is described in EP 206,783. 20 µg of pYSV201 plasmid was digested with BamHI and partially digested with PstI. To this partial digest was added the following oligonucleotides:

5' AATTCGATGAGATTTTCCTTCAATTTTTACTGCA 3'
3' GCTACTCTAAAGGAAGTTAAAAATG 5'.

5 Only the antisense strand of the oligonucleotide was
kinase labelled so that the oligonucleotides did not
polymerize at the 5'-end. After acrylamide gel
electrophoresis (10%), the fragment of 385 bp was
isolated by electroelution. This EcoRI- BamHI fragment
of 385 bp was cloned into pA0202 which had been cut with
EcoRI and BamHI. Routinely, 5 ng of vector cut with the
10 appropriate enzymes and treated with calf intestine
alkaline phosphatase, was ligated with 50 ng of the
insert fragment. MC1061 cells were transformed, amp^r
cells were selected, and DNA was prepared. In this case,
the resulting plasmid, pA0203, was cut with EcoRI and
15 BglIII to yield a fragment of greater than 700 bp. The α -
factor-GRF fragment codes for the (1-40)leu²⁷ version of
GRF and contains the processing sites lys-arg-glu-ala-
glu-ala.

20 The AOX1 promoter was isolated as a 1900 bp
EcoRI fragment from 20 μ g of pAOP3 and subcloned into
EcoRI-digested pA0203. The development of pAOP3 is
disclosed in EP 226,846 and described hereinbelow.
MC1061 cells were transformed with the ligation reaction,
amp^r colonies were selected, and DNA was prepared. The
25 correct orientation contains a \approx 376 bp HindIII fragment,
whereas the wrong orientation has an \approx 675 bp fragment.
One such transformant was isolated and was called pA0204.

The parent vector for pA0208 is the HIS4, PARS2
plasmid pYJ32 (NRRL B-15891) which was modified to change
30 the EcoRV site in the tet^R gene to a BglIII site, by
digesting pYJ32 with EcoRV and adding BglII linkers to
create pYJ32(+BglIII). This plasmid was digested with
BglII and the 1.75 Kb BglIII fragment from pA0204
containing the AOX1 promoter- α factor GRF-AOX1 3'
35 expression cassette was inserted. The resulting vector
was called pA0208. The orientation was verified by an

EcoRI digest yielding an 850 bp fragment + vector, as opposed to 1.1 Kb + vector in the other orientation.

a. Construction of plasmid pAOP3:

- 5 1. Plasmid pPG2.5 [a pBR322 based plasmid containing the approximately 2.5 Kbp EcoRI-SalI fragment from plasmid pPG4.0, which plasmid contains the primary alcohol oxidase gene (AOX1) and regulatory regions and which is available in an E. coli host from the Northern
10 Regional Research Center of the United States Department of Agriculture in Peoria, Illinois as NRRL B-15868] was linearized with BamHI.
2. The linearized plasmid was digested with BAL31;
3. The resulting DNA was treated with E. coli
15 DNA Polymerase I Klenow fragment to enhance blunt ends, and ligated to EcoRI linkers;
4. The ligation products were transformed into E. coli strain MM294;
5. Transformants were screened by the colony
20 hybridization technique using a synthetic oligonucleotide having the following sequence:
5'TTATTCGAAACGGAATTCC.
This oligonucleotide contains the AOX1 promoter sequence up to, but not including, the ATG initiation codon, fused
25 to the sequence of the EcoRI linker;
6. Positive clones were sequenced by the Maxam-Gilbert technique. All three positives had the following sequence:
30 5'...TTATTCGAAACGAGGAATTCC...3'.
They all retained the "A" of the ATG (underlined in the above sequence). It was decided that this A would probably not be detrimental; thus all subsequent clones are derivatives of these positive clones. These clones have been given the laboratory designation pAOP1, pAOP2
35 and pAOP3 respectively.

III. Construction of plasmid pA0815:

Plasmid pA0815 was constructed by mutagenizing plasmid pA0807 (described hereinbelow) to change the ClaI site downstream of the AOX1 transcription terminator in pA0807 to a BamHI site. The oligonucleotide used for mutagenizing pA0807 had the following sequence: 5' GAC GTT CGT TTG TGC GGA TCC AAT GCG GTA GTT TAT 3'. The mutagenized plasmid was called pA0807-Bam. Plasmid pA0804 was digested with BglII and 25 ng of the 2400 bp fragment were ligated to 250 ng of the 5400 bp BglII fragment from BglII-digested pA0807-Bam. The ligation mix was transformed into MC1061 cells and the correct construct was verified by digestion with Pst/BamHI to identify 5700 and 2100 bp sized bands. The correct construct was called pA0815. The restriction map of the expression vector pA0815 is shown in Figure 5.

a. Plasmid pA0807 was constructed as follows:

1. Preparation of fl-ori DNA

fl bacteriophage DNA (50 μ g) was digested with 50 units of Rsa I and Dra I (according to manufacturer's directions) to release the \approx 458 bp DNA fragment containing the fl origin of replication (ori). The digestion mixture was extracted with an equal volume of phenol: chloroform (V/V) followed by extracting the aqueous layer with an equal volume of chloroform and finally the DNA in the aqueous phase was precipitated by adjusting the NaCl concentration to 0.2M and adding 2.5 volumes of absolute ethanol. The mixture was allowed to stand on ice (4°C) for 10 minutes and the DNA precipitate was collected by centrifugation for 30 minutes at 10,000 x g in a microfuge at 4°C.

The DNA pellet was washed 2 times with 70% aqueous ethanol. The washed pellet was vacuum dried and dissolved in 25 μ l of TE buffer [1.0 mM EDTA in 0.01 M (pH7.4) Tris buffer]. This DNA was electrophoresed on 1.5% agarose gel and the \approx 458 bp fl-ori fragment was electroeluted onto DE81 (Whatman) paper and eluted from

the paper in 1M NaCl. The DNA solution was precipitated as detailed above and the DNA precipitate was dissolved in 25 μ l of TE buffer (fl-ori fragment).

2. Cloning of fl-ori into Dra I sites of

5 pBR322

pBR322 (2 μ g) was partially digested with 2 units Dra I (according to manufacturer's instructions). The reaction was terminated by phenol:chloroform extraction followed by precipitation of DNA as detailed in step 1 above. The DNA pellet was dissolved in 20 μ l of TE buffer. About 100 ng of this DNA was ligated with 100 ng of fl-ori fragment (step 1) in 20 μ l of ligation buffer by incubating at 14°C for overnight with 1 unit of T4 DNA ligase. The ligation was terminated by heating to 15 70°C for 10 minutes and then used to transform E. coli strain JM103 [Janisch-Perron et al., Gene 22, 103 (1983)]. Amp^R transformants were pooled and superinfected with helper phage R408 [Russel et al., Supra]. Single stranded phages were isolated from the 20 media and used to reinfect JM103. Amp^R transformants contained pBRfl-ori which contains fl-ori cloned into the Dra I sites (nucleotide positions 3232 and 3251) of pBR322.

25 3. Construction of plasmid pA0807

pBRfl-ori (10 μ g) was digested for 4 hours at 37°C with 10 units each of Pst I and Nde I. The digested DNA was phenol:chloroform extracted, precipitated and dissolved in 25 μ l of TE buffer as detailed in step 1 30 above. This material was electrophoresed on a 1.2% agarose gel and the Nde I - Pst I fragment (approximately 0.8 kb) containing the fl-ori was isolated and dissolved in 20 μ l of TE buffer as detailed in step 1 above. About 100 ng of this DNA was mixed with 100 ng of pA0804 35 (described hereinafter) that had been digested with Pst I and Nde I and phosphatase-treated. This mixture was ligated in 20 μ l of ligation buffer by incubating for

overnight at 14°C with 1 unit of T4 DNA ligase. The ligation reaction was terminated by heating at 70°C for 10 minutes. This DNA was used to transform *E. coli* strain JM103 to obtain pA0807.

5 b. Plasmid pA0804 employed in the above procedure was constructed as follows:

Plasmid pBR322 was modified as follows to eliminate the EcoRI site and insert a BglII site into the PvuII site:

10 pBR322 was digested with EcoRI, the protruding ends were filled in with Klenow Fragment of E. coli DNA polymerase I, and the resulting DNA was recircularized using T4 ligase. The recircularized DNA was used to transform *E. coli* MC1061 to ampicillin-resistance and
15 transformants were screened for having a plasmid of about 4.37 kbp in size without an EcoRI site. One such transformant was selected and cultured to yield a plasmid, designated pBR322 Δ RI, which is pBR322 with the EcoRI site replaced with the sequence:

20 5'-GAATTAATTC-3'
 3'-CTTAATTAAG-5'.

pBR322 Δ RI was digested with PvuII and the linker, of sequence

25 5'-CAGATCTG-3'
 3'-GTCTAGAC-5'

was ligated to the resulting blunt ends employing T4 ligase. the resulting DNAs were recircularized, also with T4 ligase, and then digested with BglII and again
30 recircularized using T4 ligase to eliminate multiple BglII sites due to ligation of more than one linker to the PvuII-cleaved pBR322 Δ RI. The DNAs, treated to eliminate multiple BglII sites, were used to transform *E. coli* MC1061 to ampicillin-resistance. Transformants were
35 screened for a plasmid of about 4.38 kbp with a BglII site. One such transformant was selected and cultured to yield a plasmid, designated pBR322 Δ RIBGL, for further

work. Plasmid pBR322 Δ RIBGL is the same as pBR322 Δ RI except that pBR322 Δ RIBGL has the sequence.

5'-CAGCAGATCTGCTG-3'

3'-GTCGTCTAGACGAC-5'

5 in place of the PvuII site in pBR322 Δ RI.

pBR322 Δ RIBGL was digested with a SalI and BglII and the large fragment (approximately 2.97 kbp) was isolated. Plasmid pBSAGI5I, which is described in European Patent Application Publication No. 0,226,752, was digested completely with BglII and XhoI and an approximately 850 bp fragment from a region of the P. pastoris AOX1 locus downstream from the AOX1 gene transcription terminator (relative to the direction of transcription from the AOX1 promoter) was isolated. The BglII-XhoI fragment from pBSAGI5I and the approximately 2.97 kbp, SalI-BglII fragment from pBR322 Δ RIBGL were combined and subjected to ligation with T4 ligase. The ligation mixture was used to transform E. coli MC1061 cells to ampicillin-resistance and transformants were screened for a plasmid of the expected size (approximately 3.8 kbp) with a BglII site. This plasmid was designated pA0801. The overhanging end of the SalI site from the pBR322 Δ RIBGL fragment was ligated to the overhanging end of the XhoI site on the 850 bp pBSAGI5I fragment and, in the process, both the SalI site and the XhoI site in pA0801 were eliminated.

pBSAGI5I was then digested with ClaI and the approximately 2.0 kbp fragment was isolated. The 2.0 kbp fragment has an approximately 1.0-kbp segment which comprises the P. pastoris AOX1 promoter and transcription initiation site, an approximately 700 bp segment encoding the hepatitis B virus surface antigen ("HBsAg") and an approximately 300 bp segment which comprises the P. pastoris AOX1 gene polyadenylation signal and site-encoding segments and transcription terminator. The HBsAg coding segment of the 2.0 kbp fragment is terminated, at the end adjacent the 1.0 kbp segment with

the AOX1 promoter, with an EcoRI site and, at the end adjacent the 300 bp segment with the AOX1 transcription terminator with a StuI site, and has its subsegment which codes for HBsAg oriented and positioned, with respect to the 1.0 kbp promoter-containing and 300 bp transcription terminator-containing segments, operatively for expression of the HBsAg upon transcription from the AOX1 promoter. The EcoRI site joining the promoter segment to the HBsAg coding segment occurs just upstream (with respect to the direction of transcription from the AOX1 promoter) from the translation initiation signal-encoding triplet of the AOX1 promoter.

For more details on the promoter and terminator segments of the 2.0 kbp, ClaI-site-terminated fragment of pBSAGI5I, see European Patent Application Publication No. 0,226,846 and Ellis *et al.*, Mol. Cell Biol. 5, 1111 (1985).

Plasmid pA0801 was cut with ClaI and combined for ligation using T4 ligase with the approximately 2.0 kbp ClaI-site-terminated fragment from pBSAGI5I. The ligation mixture was used to transform E. coli MC1061 to ampicillin resistance, and transformants were screened for a plasmid of the expected size (approximately 5.8 kbp) which, on digestion with ClaI and BglII, yielded fragments of about 2.32 kbp (with the origin of replication and ampicillin-resistance gene from pBR322) and about 1.9 kbp, 1.48 kbp, and 100 bp. On digestion with BglII and EcoRI, the plasmid yielded an approximately 2.48 kbp fragment with the 300 bp terminator segment from the AOX1 gene and the HBsAg coding segment, a fragment of about 900 bp containing the segment from upstream of the AOX1 protein encoding segment of the AOX1 gene in the AOX1 locus, and a fragment of about 2.42 kbp containing the origin of replication and ampicillin resistance gene from pBR322 and an approximately 100 bp ClaI-BglII segment of the AOX1 locus (further upstream from the AOX1-encoding

segment than the first mentioned 900 bp EcoRI-BglII segment). Such a plasmid had the ClaI fragment from pBSAGI5I in the desired orientation, in the opposite undesired orientation, there would be EcoRI-BglII fragments of about 3.3 kbp, 2.38 kbp and 900 bp.

One of the transformants harboring the desired plasmid, designated pA0802, was selected for further work and was cultured to yield that plasmid. The desired orientation of the ClaI fragment from pBSAGI5I in pA0802 had the AOX1 gene in the AOX1 locus oriented correctly to lead to the correct integration into the P. pastoris genome at the AOX1 locus of linearized plasmid made by cutting at the BglII site at the terminus of the 800 bp fragment from downstream of the AOX1 gene in the AOX1 locus.

pA0802 was then treated to remove the HBsAg coding segment terminated with an EcoRI site and a StuI site. The plasmid was digested with StuI and a linker of sequence:

5'-GGAATTCC-3'

3'-CCTTAAGG-5'

was ligated to the blunt ends using T4 ligase. The mixture was then treated with EcoRI and again subjected to ligating using T4 ligase. The ligation mixture was then used to transform E. coli MC1061 cells to ampicillin resistance and transformants were screened for a plasmid of the expected size (5.1 kbp) with EcoRI-BglII fragments of about 1.78 kbp, 900 bp, and 2.42 kbp and BglII-ClaI fragment of about 100 bp, 2.32 kbp, 1.48 kbp, and 1.2 kbp. This plasmid was designated pA0803. A transformant with the desired plasmid was selected for further work and was cultured to yield pA0803.

Plasmid pA0804 was then made from pA0803 by inserting, into the BamHI site from pBR322 in pA0803, an approximately 2.75 kbp BglII fragment from the P. pastoris HIS4 gene. See, e.g., Cregg et al., Mol.

Cell. Biol. 5, 3376 (1985) and European Patent Application Publication Nos. 0,180,899 and 0,188,677. pA0803 was digested with BamHI and combined with the HIS4 gene-containing BglIII site-terminated fragment and the mixture subjected to ligation using T4 ligase. The ligation mixture was used to transform E. coli MC1061 cells to ampicillin-resistance and transformants were screened for a plasmid of the expected size (7.85 kbp), which is cut by SalI. One such transformant was selected for further work, and the plasmid it harbors was designated pA0804.

pA0804 has one SalI-ClaI fragment of about 1.5 kbp and another of about 5.0 kbp and a ClaI-ClaI fragment of 1.3 kbp; this indicates that the direction of transcription of the HIS4 gene in the plasmid is the same as the direction of transcription of the ampicillin resistance gene and opposite the direction of transcription from the AOX1 promoter.

The orientation of the HIS4 gene in pA0804 is not critical to the function of the plasmid or of its derivatives with cDNA coding segments inserted at the EcoRI site between the AOX1 promoter and terminator segments. Thus, a plasmid with the HIS4 gene in the orientation opposite that of the HIS4 gene in pA0804 would also be effective for use in accordance with the present invention.

Example 2

Strain development and characterization

Plasmid pSCD103, the construction of which is described in Example 1, was used to develop Mut⁺ and Mut⁻ strains of P. pastoris. The His⁻ strain GS115 (ATCC 20864) was the host for all transformations. Transformations were accomplished by the whole-cell LiCl method [Ito et al., J. Bacteriol. 153(1), 163 (1983)], with minor modification necessary for adaptation to P. pastoris.

To develop Mut⁺ strains, pSCD103 was digested with SacI, which linearizes the vector within the AOX1 promoter region, and 10 µg of the linearized vector were used to transform GS115. Histidine prototrophs were selected.

5

To develop Mut⁻ strains, pSCD103 was digested with BglIII thereby liberating an expression cassette comprised of the AOX1 promoter region, α MF leader-V, gene, AOX1 transcription termination signals, HIS4 gene for selection, and AOX1 3' region. Both ends of this expression cassette contain long sequences which are homologous to the 5' and 3' ends of the AOX1 locus. 10 µg of the linearized vector were used to transform GS115 cells. Histidine prototrophs were selected and screened for the Mut⁻ phenotype by replica plating colonies from glucose containing media to methanol containing media, and evaluating growth rate on methanol. Slow growth on methanol was indicative of the Mut⁻ phenotype. Several His⁺Mut⁻ colonies were identified.

10

15

20

To characterize the Mut⁺ and Mut⁻ transformants for cassette copy number and site of integration, DNA from several of the selected colonies was digested with EcoRI and probed with nick-translated pSCD103. The Southern analysis yielded the following information:

25

<u>Strain name</u>	<u>Mut[±]</u>	<u>Copy</u>	<u>Site of</u>
		<u>Number</u>	<u>Integration</u>
G+SCD103S03	Mut ⁺	one	<u>AOX1</u>
G+SCD103S16	Mut ⁺	two	<u>AOX1</u>
30 G-SCD103S03	Mut ⁻	one	<u>AOX1</u> (disruption)

30

Example 3

Fermentation in two-liter fermentors

a. Fermentor start-up and general operation

35

The 2-liter fermentors (Biolafitte, ISL Biolafitte, Princeton, NJ) were autoclaved at a volume of one liter containing 5X Basal Salts (21 ml/l 85%

phosphoric acid, 0.9 g/l Calcium Sulfate x 2H₂O, 14.3 g/l Potassium Sulfate, 11.7 g/l Magnesium Sulfate x 7H₂O, 3.25 g/l Potassium Hydroxide) and 5% (w/v) glycerol.

After sterilization, 5 ml of a PTM₁ trace salts solution
5 (6.0 g/l Cupric Sulfate x 5H₂O, 0.08 g/l Sodium Iodide, 3.0 g/l Manganese Sulfate x H₂O, 0.2 g/l Sodium Molybdate x 2H₂O, 0.02 g/l Boric Acid, 0.5 g/l Cobalt Chloride, 20 g/l Zinc Chloride, 65 g/l Ferrous Sulfate x 7H₂O, 0.2 g/l Biotin, and 5.0 ml of Sulfuric Acid) was added, and the
10 pH of the fermentor was adjusted to 3.0 with the addition of concentrated Ammonium Hydroxide. During the fermentation, the pH was controlled at 3.0 with the addition of a 50% (v/v) Ammonium Hydroxide solution.

The fermentors were then inoculated with 50 ml
15 of an overnight culture grown in 6.75 g/l Difco yeast nitrogen base, 2% glycerol, 0.1 M potassium phosphate, pH 6.0. After 16 hours of fermentor growth, the pH of the medium was dropped to 2.6 and the cells continued to grow in a batch mode to exhaust the original charge of
20 glycerol. Upon glycerol exhaustion, a 50% (w/v) glycerol feed containing 12 ml/l PTM₁ trace salts was initiated at a feed rate of 5 to 20 ml/h. After 200 ml of the glycerol feed was added into the fermentor, a 100 % methanol feed containing 12 ml/l PTM₁ trace salts, was
25 initiated at 1 ml/h, and the glycerol feed was shut off after 1 hour of methanol feeding. After 4 hours of methanol feeding, the methanol feed was increased to 5-6 ml/h over an 8-12 hour period and was maintained at this rate for the remainder of the fermentation. The
30 dissolved oxygen concentration was maintained above 20% of saturation by adjusting agitation and aeration as needed. The temperature was controlled at 30°C and foaming was controlled by the addition of a 5% solution of Struktol J-673 antifoam
35 (Struktol Co., Stow, OH).

Before harvesting the fermentor, the pH was decreased to 2.5 with the addition of 85 % phosphoric

acid. The contents were then centrifuged to remove cells and the supernatant was filter-sterilized through a 0.22 μ Corning filter (Corning Glass Co., Corning, NY). The supernatant was then frozen at -20°C.

5

b. Growth of Mut⁺ and Mut⁻ strains

Run 568: G+SCD103S03

Run 570: G+SCD103S16

Run 571: G-SCD103S03

Run 585: G+SCD103S03

10

Run 593: G+SCD103S16

Fermentation Runs 568, 570, 571, 585, and 593 were conducted as described above, except that Runs 568, 570, and 571 were conducted at pH 5.0; Run 585 was performed at pH 3.5 and the pH was not adjusted to pH 2.5 at the end of the fermentation run; Run 593 was conducted as hereinabove described.

15

Figure 6 shows the time course for cell yield for one-liter fermentation runs with strains G+SCD103S03 (Run 568), G+SCD103S16 (Run 570), and G-SCD103S03 (Run 20 571). Cell yield was calculated as the mass of wet cells per liter of broth after centrifugation. A conversion factor of 0.25 was used to calculate yield of dry cells per liter.

20

The single-copy Mut⁺ and Mut⁻ strains grew at 25 equivalent rates, whereas the two-copy Mut⁺ strain showed slightly decreased cell yield on methanol. However, because Pichia transformants carrying multiple copies of an expression cassette may express higher levels of heterologous proteins in the fermentor than do the 30 strains with a single-copy, another fermentation, Run 593 was conducted to analyze the level of recombinant CD-V₁ in the broth of the two-copy strain, G+SCD103S16.

30

Figures 7A and 7B show the time course for cell 35 yield and CD4-V₁ production, respectively, for fermentation Run 593. The expression level was determined for unfiltered, reduced broth samples, and was estimated by quantitative Western blot analysis to be 130

35

mg/liter after 71 hours on methanol. The level was continuing to increase when the fermentor was harvested.

Example 4

5 Analysis of secreted CD4-V₁

a. Western blot analysis

The V₁ region of CD4 contains a single disulfide bond between two cysteine residues, which are located at positions 16 and 84 of the mature CD4, near the N- and C-termini, respectively. Therefore, non-reduced V₁ molecules from fermentor broth samples will co-migrate with the V₁ standard, regardless of whether the molecule has been nicked between the cysteines. On the other hand, reduced samples will only co-migrate with the standard if the peptide bonds between the cysteines are intact. Separating *P. pastoris* broth samples on non-reducing gels yielded a quantitative measurement of the total amount of V₁ contained in the fermentor broth, while reducing gels yielded the amount of intact V₁.

20 Fermentor broth samples from Runs 571 (Mut⁻), 568 and 585 (single copy Mut⁺) and 593 (multi-copy Mut⁺) were analyzed by Western blotting. Ten microliters of each sample were mixed with an equal volume of 2x Laemmli sample buffer containing 200 mM DTT (+DTT). In some cases, the 2x sample buffer lacked any reducing agent (-DTT). Fermentor samples and 2X sample buffer were mixed and immediately boiled for 5 minutes (+DTT), or mixed and immediately placed at room temperature until the gel was loaded (-DTT). Samples thus prepared were separated by electrophoresis, at 4°C, on 15% SDS-PAGE gels, at 150V constant voltage, until the bromophenol blue tracking dye had reached the bottom of the gel.

30 *E. coli* produced V₁ (Smith Kline & French), used as standard, was treated in an identical manner, and separated on the same gels as a standard control. Reduced (+DTT) and non-reduced (-DTT) samples were separated on different gels. For quantitation of the

P. pastoris produced V₁ in fermentor broth, fermentor samples were separated in non-adjacent lanes, to prevent spillover errors, and several different amounts of V₁ standards were separated on the same gel to generate an internal standard curve.

Gels were transblotted to nitrocellulose (0.1 μ pore size) for 90 minutes at 4°C, using a carbonate buffer system [S.D.Dunn, Anal. Biochem. 157, 144 (1986)]. The filters were blocked for 16 hours at room temperature in Western blocking buffer (WBB), incubated for two hours with a 1:1000 dilution of rabbit anti-sCD4 [SK&F; Arthos et al., Cell 57, 469 (1989)] in WBB at room temperature, washed four-times for 15 minutes in WBB, incubated for one hour at room temperature in a 1:5000 dilution of low specific activity ¹²⁵I-Protein A (New England Nuclear) in WBB, washed four-times for 15 minutes each time in WBB, air dried and exposed to Kodak X-omat film at -70°C, with two intensifying screens. V₁ bands, identified by reaction with anti-sCD4, were excised from the nitrocellulose filters and quantitated using a gamma-counter. The fermentor samples were quantitated by comparison with the V₁ standard curve on each filter. The results of these analyses are summarized in the following Table:

			<u>V₁ YIELD (mg/L)</u>		
<u>RUN</u>	<u>MUT+/-</u>	<u>pH</u>	<u>TOTAL</u>	<u>INTACT</u>	<u>%INTACT</u>
571	Mut ⁻	5.0	100	0	0
568	single copy Mut ⁺	5.0	100	0	0
585	single copy Mut ⁺	3.5	125	30	25
570	two-copy Mut ⁺	5.0	100	0	0
593	two-copy Mut ⁺	2.6	100	100	100

b. Amino Acid Sequencing

We have further characterized the Pichia-produced recombinant CD4-V₁ by determining the N-terminal sequence of the protein which comigrates with the V₁ standard on reducing SDS-PAGE. As shown in the silver-

stained reducing gel of V_1 standard and fermentor broth samples pictured in Figure 8, control fermentor broth obtained from the fermentation of Pichia strain G-PA0815 does not contain a protein which comigrates with the V_1 standard. In contrast, Pichia rCD4- V_1 appears to be the major lower molecular weight protein species present in broth samples from the fermentation of strain G+SCD103S16. To ensure that broth components did not affect the migration or staining of V_1 in polyacrylamide gels, E. coli-derived V_1 standard and rCD4- V_1 -containing broth from fermentation of G+SCD103S16 were separately mixed with Pichia control broth and analyzed by reducing SDS-PAGE. As shown in Figure 8, the electrophoretic characteristics of the V_1 standard and Pichia-produced rCD4- V_1 were unaltered by exposure to Pichia control broth; the V_1 standard and Pichia rCD4- V_1 co-migrated to the same gel position and exhibited similar staining properties in the presence and absence of control broth. The first 15 residues of Pichia rCD4- V_1 were determined, and found to be identical to the published sequence (Maddon et al., Supra) for mature human CD4- V_1 (Figure 9). From this result, it was concluded that the N-terminus of the recombinant CD4- V_1 was correctly processed from the α MF leader.

c. Stability

The stability of the V_1 molecules in the fermentor broth, which had been adjusted to pH 2.5, was analyzed in two ways. First, the stability of V_1 during storage was analyzed by subjecting identical broth samples to a freeze-thaw cycle, followed by incubation of the samples for 20 hours, under varying conditions (Figure 10, lanes 2-4). No change was observed in the amount of immunoreactive material or the proportion of intact V_1 in the different samples. In the second experiment, the pH of the broth samples was raised to 5.0 and the samples incubated under the same conditions as before (Figure 10, lanes 5-7). As seen in the Figure,

some broadening of the intact V₁ band occurred under these conditions. Therefore, while the rCD4-V₁ is stable in the pH 2.5 broth samples, some degree of proteolytic degradation may occur in samples at elevated pHs.

CLAIMS:

1. A Pichia pastoris (P. pastoris) cell containing in its genome at least one copy of a DNA sequence operably encoding in P. pastoris at least a
5 portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the human immunodeficiency virus (HIV), in operational association with a DNA sequence encoding a signal sequence which functions to direct secretion of said human CD4 glycoprotein or a
10 portion thereof in P. pastoris, both under the regulation of a promoter region of a P. pastoris gene.
2. A P. pastoris cell according to Claim 1, wherein said signal sequence-encoding DNA comprises a DNA sequence encoding the S. cerevisiae AMF pre-pro sequence,
15 and a DNA sequence encoding AMF processing-site lys-arg.
3. A P. pastoris cell according to Claim 2, wherein said P. pastoris gene is the P. pastoris AOX1 gene.
4. A P. pastoris cell according to Claim 3
20 wherein said DNA sequence operably encodes in P. pastoris the V₁ region of human CD4 glycoprotein.
5. A P. pastoris cell according to Claim 4 containing at least two copies of said DNA sequences.
6. A P. pastoris cell containing in its
25 genome at least one copy of an expression cassette comprising in the direction of transcription, a promoter region of a first P. pastoris gene, a DNA sequence operably encoding in P. pastoris at least a portion of human CD4 glycoprotein, containing the site of
30 interaction between CD4 and the HIV virus, preceeded by a DNA sequence encoding a signal sequence directing the secretion of said glycoprotein or a portion thereof in P. pastoris, and a transcription terminator of a second P. pastoris gene, said first and second P. pastoris genes
35 being identical or different, the segments of said expression cassette being in operational association.

7. A P. pastoris cell according to Claim 6, wherein said signal sequence-encoding DNA comprises a DNA sequence encoding the S. cerevisiae AMF pre-pro sequence and a DNA sequence encoding AMF processing-site lys-arg.

5

8. A P. pastoris cell according to Claim 7 wherein said first and second P. pastoris genes are identical and are the P. pastoris AOX1 gene.

10

9. A P. pastoris cell according to Claim 8 wherein said DNA sequence operably encodes in P. pastoris the V₁ region of human CD4 glycoprotein.

10. A P. pastoris cell according to Claim 9 containing at least two copies of said expression cassette.

15

11. A P. pastoris cell according to Claim 10, containing two copies of said expression cassette integrated by addition at the AOX1 locus of said P. pastoris genome.

20

12. A P. pastoris cell according to Claim 9, containing a single copy of said expression cassette integrated by addition at the AOX1 locus of said P. pastoris genome.

25

13. A P. pastoris cell according to Claim 9, containing a single copy of said expression cassette integrated by gene replacement at the AOX1 locus of said P. pastoris genome.

30

14. A DNA fragment optionally contained within, or which is, a circular plasmid comprising at least one copy of an expression cassette comprising in the direction of transcription, a promoter region of a first P. pastoris gene, a DNA sequence operably encoding in P. pastoris at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the HIV virus, preceded by a DNA sequence encoding a signal sequence directing the secretion of said glycoprotein or a portion thereof in P. pastoris, and a transcription terminator of a second P. pastoris gene, said first and second P. pastoris genes being

35

identical or different, the segments of said expression cassette being in operational association.

15. A DNA fragment according to Claim 14, wherein said signal sequence-encoding DNA comprises a DNA sequence encoding the S. cerevisiae AMF pre-pro sequence, and a DNA sequence encoding AMF processing-site lys-arg.

16. A DNA fragment according to Claim 15, wherein said first and second P. pastoris genes are identical and are the P. pastoris AOX1 gene.

17. A DNA fragment according to Claim 16 wherein said DNA sequence operably encodes in P. pastoris the V₁ region of human CD4 glycoprotein.

18. A DNA fragment according to Claim 16, further comprising a selectable marker gene and ends having sufficient homology with a target gene to effect integration of said DNA fragment therein.

19. A DNA fragment according to Claim 18, wherein said target gene is the P. pastoris AOX1 gene.

20. A DNA fragment according to Claim 18 which is a BglIII digest of the expression vector pSCD103.

21. A DNA fragment according to Claim 18, which is a SacI digest of the expression vector pSCD103.

22. An expression vector containing at least one copy of an expression cassette comprising in the direction of transcription, a promoter region of a first P. pastoris gene, a DNA sequence operably encoding in P. pastoris at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the HIV virus, preceded by a DNA sequence encoding a signal sequence directing the secretion of said glycoprotein or a portion thereof in P. pastoris, and a transcription terminator of a second P. pastoris gene, said first and second P. pastoris genes being identical or different, the segments of said expression cassette being in operational association.

23. An expression vector according to Claim 22, wherein said signal sequence-encoding DNA comprises a DNA sequence encoding the S. cerevisiae AMF pre-pro sequence, and a DNA sequence encoding AMF processing-site lys-arg.

24. An expression vector according to Claim 23, further comprising sequences allowing for its replication and selection in bacteria.

25. An expression vector according to Claim 24, which is a pBR322 derivative.

26. An expression vector according to Claim 25, which is the Pichia expression vector pSCD103.

27. A culture of viable P. pastoris cells according to any one of Claims 1 to 13.

28. A process for producing and secreting at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the HIV virus, into the culture medium comprising growing P. pastoris transformants containing in their genome at least one copy of a DNA sequence operably encoding in P. pastoris at least a portion of human CD4 glycoprotein, containing the site of interaction between CD4 and the HIV virus, in operational association with a DNA sequence encoding a signal sequence directing the secretion of said glycoprotein or a portion thereof in P. pastoris, both under the regulation of a promoter region of a P. pastoris gene, under conditions allowing the expression of said DNA sequences in said P. pastoris and secretion of said glycoprotein or a protein thereof into the culture medium in a substantially pure form, substantially devoid of degradation products.

29. A process according to Claim 28, wherein said signal sequence is the S. cerevisiae AMF pre-pro sequence.

30. A process according to Claim 29, wherein said transformants are developed from the P. pastoris his4⁻ strain GS115.

31. A process according to Claim 30, wherein said transformants have the Mut⁺ phenotype.

32. A process according to Claim 28, which comprises:

5 a. growing said P. pastoris transformants on a medium containing repressing carbon source to generate cell mass in absence of heterologous gene expression,

 b. continuing growth under glycerol limitation conditions, and

10 c. initiating heterologous gene expression by adding methanol to the medium, and keeping the pH at or below about 3.5 during said heterologous gene expression.

 33. A process according to Claim 32, wherein the pH is kept between about 2.5 and about 3.5 during
15 heterologous gene expression.

 34. A process according to Claim 33, wherein the pH is kept between about 2.5 and about 3.0 during heterologous gene expression.

20 35. A process according to any one of Claims 28 to 34, further comprising the step of harvesting said human CD4 glycoprotein or a portion thereof from the culture medium.

25 36. A process for producing a heterologous protein in P. pastoris, wherein the pH of the culture medium is maintained at or below about 3.5 during heterologous gene expression.

 37. A process according to Claim 36, wherein said heterologous protein is secreted into the fermentation medium.

30 38. A process according to Claim 36, wherein the pH is maintained between about 2.5 and about 3.5.

 39. Substantially pure human CD4 glycoprotein or a portion thereof containing the site of interaction between CD4 and the human immunodeficiency virus (HIV)
35 produced in yeast.

 40. Substantially pure human CD4 glycoprotein or a portion thereof according to Claim 39 produced in P. pastoris.

1 / 15

1 ATGAGATTTCCCTTCAATTTTACTGCAGTTTATTCCGAGCATCCTCCGCAATTAGCTGCT 60
 +-----+-----+-----+-----+-----+-----+
 TACTCTAAAGGAAGTTAAATAATGACGTCAAAATAAGCGTCGTAGGAGCGGTAATCGACCGA
MetArgPheProSerIlePheThrAlaValLeuPheAlaAlaSerSerAlaLeuAlaAla

61 CCAGTCAACACTACAACAGAAGATGAACCGGCACAAATTCGGGTGAAGCTGTATCGGT 120
 +-----+-----+-----+-----+-----+-----+
 GGTCAGTTGTGATGTTGTCTTCTACTTTTCCCGTGTTTAAGGCCGACTTCGACAGTAGCCA
ProValAsnThrThrThrGluAspGluThrAlaGlnIleProAlaGluAlaValIleGly

121 TACTCAGATTTAGAAGGGGATTTCCGATGTTGCTGTTTGGCCATTTTCCAAACAGCACAAT 180
 +-----+-----+-----+-----+-----+-----+
 ATGAGTCTAAATCTCCCTAAAGCTACACGACACAAACGGTAAAGGTTGTCGTGTTTA
TyrSerAspLeuGluGlyAspPheAspValAlaValLeuProPheSerAsnSerThrAsn

181 AACGGGTTATTGTTATAAATACTACTATTGCCAGCATTTGCTGCTAAAGAAGAAGGGGTA 240
 +-----+-----+-----+-----+-----+-----+
 TTGCCCAATAACAAATATTTATGATGATAACGGTCGTAAACGACGATTTCTTCTTCCCCAT
AsnGlyLeuLeuPheIleAsnThrThrIleAlaSerIleAlaAlaLysGluGlyVal

241 TCTTTGGATAAA 252
 +-----+-----
 AGAAACCTATTT
SerLeuAspLys

FIG. 1

241
 AAAA CTCCA ACCAGATAAGATTCTGGGAAATCAGGGCTCCTTCTTAACTAAAGGTCCAT
 -----+-----+-----+-----+-----+-----+
 TTTT GAGGTTGGTCTATTTCTAAGACCCCTTTAGTCGCCGAGGAAGAATTGATTTCCAGGTA
 ysAsnSerAsnGlnIleLysIleLeuGlyAsnGlnGlySerPheLeuThrLysGlyProS
 300

301 CCAAGCTGAATGATCGCGGTGACTCAAGAAGACCTTTGGACCAAGGAAACTTCCCGC
 -----+-----+-----+-----+-----+-----+
 GGTTCGACTTACTAGCGCGACTGAGTCTCTTCGGAAACCCCTGTTCTTGAAGGGG
 erLysLeuAsnAspArgAlaAspSerArgSerLeuTrpAspGlnGlyAsnPheProL
 360

361
TGATCATCAAGAATCTTTAAGATAGAAGACTCAGATACCTTACATCTGTGAAGTGGAGGACC
-----+-----+-----+-----+-----+-----+
ACTAGTAGTCTTACAAATCTTCTGTGACTCTATGAAATGTAGACACTTCACCTCCTGG
euIIeIleLysAsnLeuLysIleGluAspSerAspThrTyrIleCysGluValGluAspG
420

421 AGAAGGAGGAGGTGCAATTGCTAGTGTTCGGATTGACTGCCAACTCGGACACCTaatcta
 -----+-----+-----+-----+-----+
 TCTTCCTCTCCAGCTTAACGATCACAAGCCTAACTGACGGTTGACGCTGTGGattagat
 lnLysGluGluValGlnLeuLeuValPheGlyLeuThrAlaAsnSerAspThr
 480

ga 481 482 ct

SUBSTITUTE SHEET

3 / 15

FIG. 2-2

361
TGATCATCAAGAACTCTTAAGATAGAAGACTCAGATACTTACATCTGTGAAGTGGAGGACC
-----+-----+-----+-----+-----+-----+
ACTAGTACTCTTAGAAATCTATCTTCTGAGTCTATGAATGTAGACACTTCACCTCCTGG
euilleLysAsnLeuLysIleGluAspSerAspThrTyrIleCysGluValGluAspG
420

421 AGAAGGAGGAGGTGCAATTGCTAGTGTTCGGATTGACTGCCAACTCGGACACCTaatcta
 -----+-----+-----+-----+-----+
 TCTTCCTCTCCAGCTTAACGATCACAAGCCTAACTGACGGTTGACGCTGTGGattagat
 lnLysGluGluValGlnLeuLeuValPheGlyLeuThrAlaAsnSerAspThr
 480

ga 481 482 ct

4 / 15

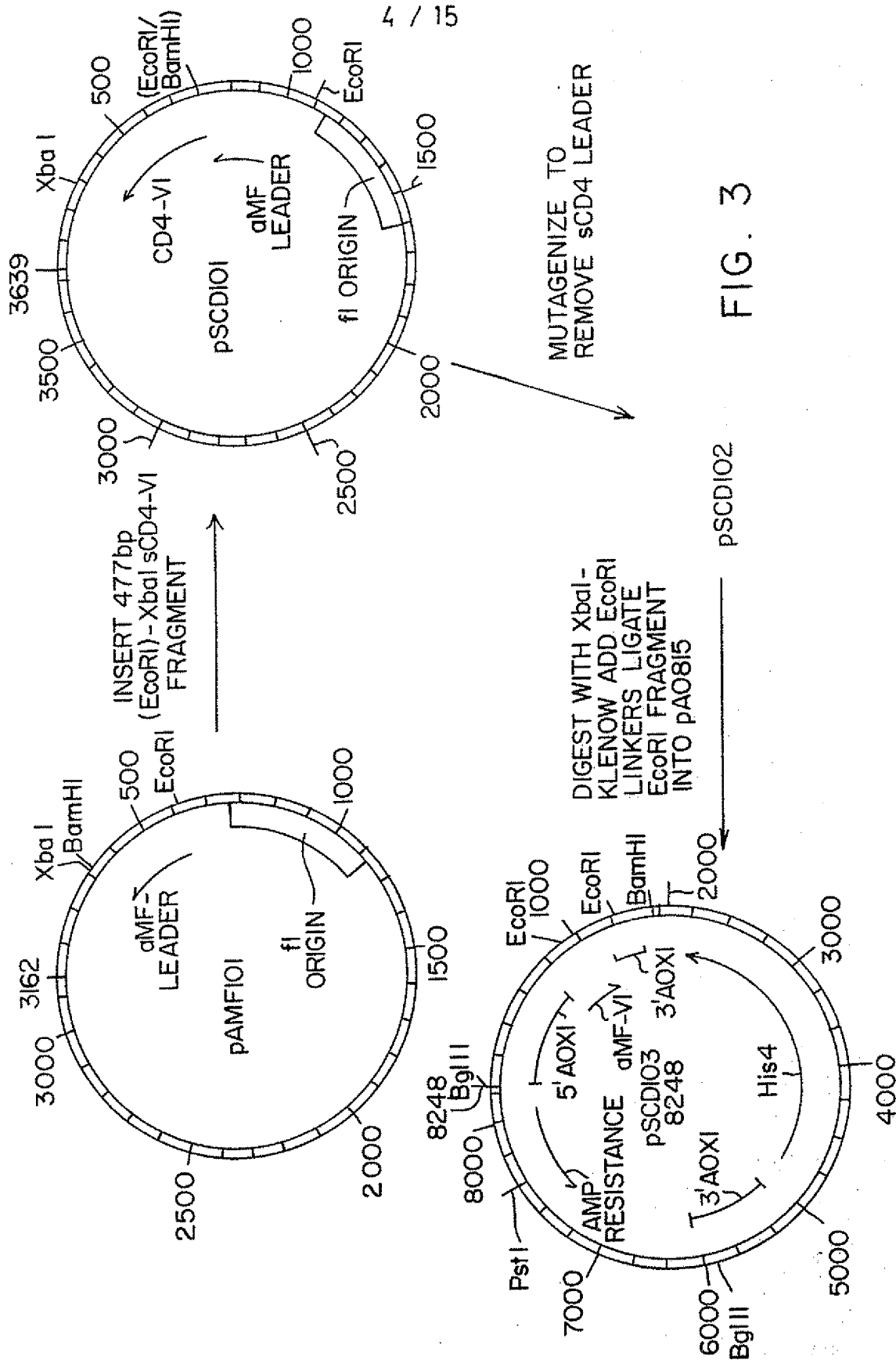


FIG. 3

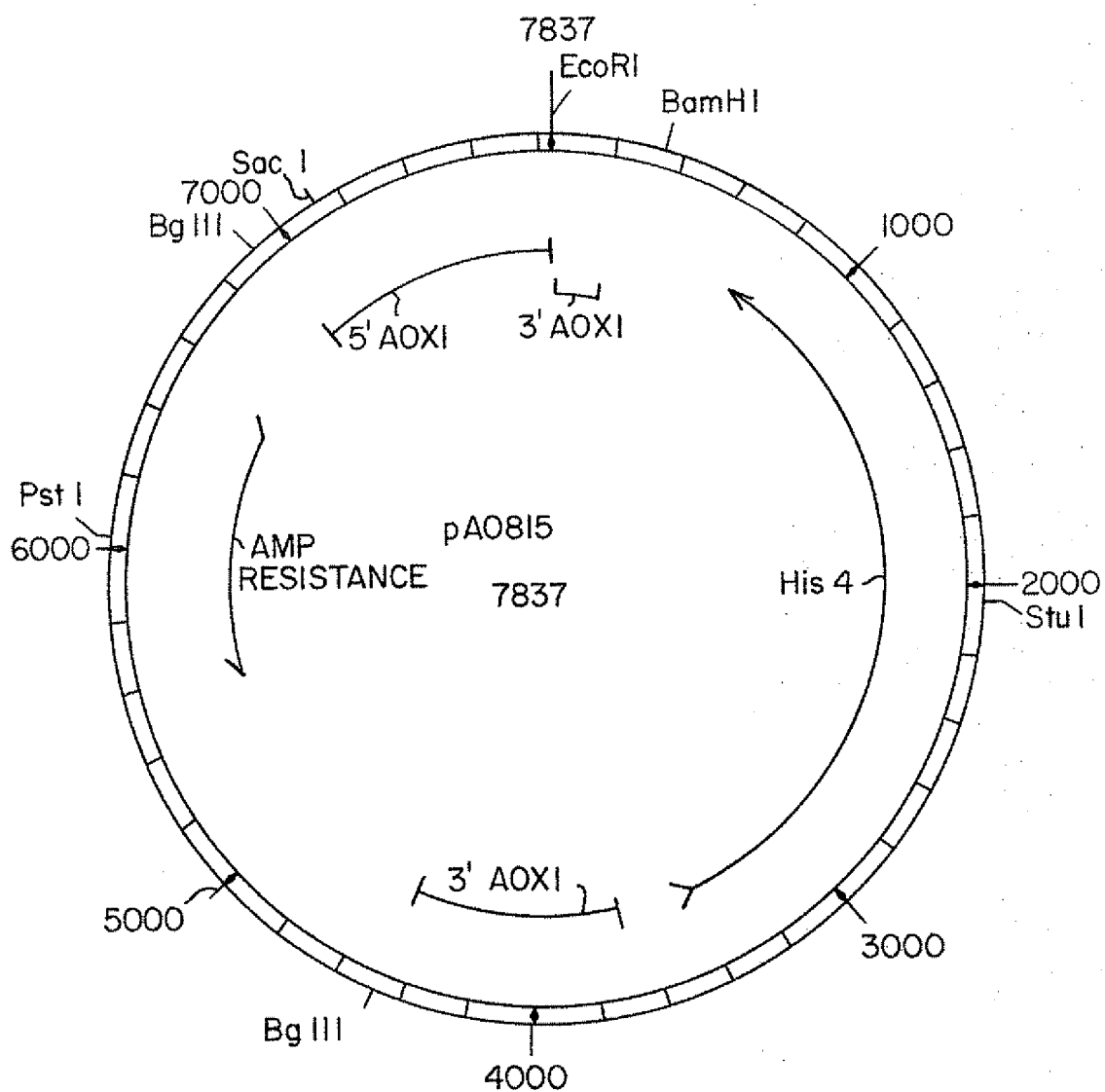
6 / 15

FIG. 4-2

301	CTGACCTGTACAGCTTCCAGAGAAGAGCATAACAATTCCACTGGAAAACTCCAACCAG -----+-----+-----+-----+-----+-----+ GACTGGACATGTCGAAGGCTCTTCTTCGTATGTTAAGGTGACCTTTTGAGGTTGGTC LeuThrCysThrAlaSerGlnLysLysSerIleGlnPheHisTrpLysAsnSerAsnGln	360
361	ATAAGATTCTGGAAATCAGGGCTCCTTCTTAATAAGGTCCATCCAAGCTGAATGAT -----+-----+-----+-----+-----+-----+ TATTTCTAAGACCCTTTAGTCCCGAGGAAGAATTGATTTCCAGTAGGTTGACTTACTA IleLysIleLeuGlyAsnGlnGlySerPheLeuThrLysGlyProSerLysLeuAsnAsp	420
421	CCCGCTCACTCAAGAAGAAGCCCTTTGGGACCAAGAAACTTCCCCCTGATCATCAAGAAT -----+-----+-----+-----+-----+-----+ GGCGGACTGAGTTCTTTCGGAAACCCCTGGTTCCTTTGAAGGGGACTAGTACTCTTA ArgAlaAspSerArgSerLeuTrpAspGlnGlyAsnPheProLeuIleLysAsn	480
481	CTTAAGATAGAAGACTCAGATACCTTACATCTGTGAAGTCGAGGACCAGAAGGAGGTG -----+-----+-----+-----+-----+-----+ GAATTCTATCTTCTGAGTCTATGAATGTAGACACTTCACCTCCTGGTCTTCCTCCTCCAC LeuLysIleGluAspSerAspThrTyrIleCysGluValGluAspGlnLysGluGluVal	540
541	CAATTGCTAGTGTTCGGATTGACTGCCAACTCGGACACCTAATCTAGGGAATTC -----+-----+-----+-----+-----+-----+ GTTAACGATCACAAGCCTAACTACGGTTGAGCCCTGTGGATTAGATCCCTTAAG GlnLeuLeuValPheGlyLeuThrAlaAsnSerAspThr	594

7 / 15

FIG. 5



SUBSTITUTE SHEET

8 / 15

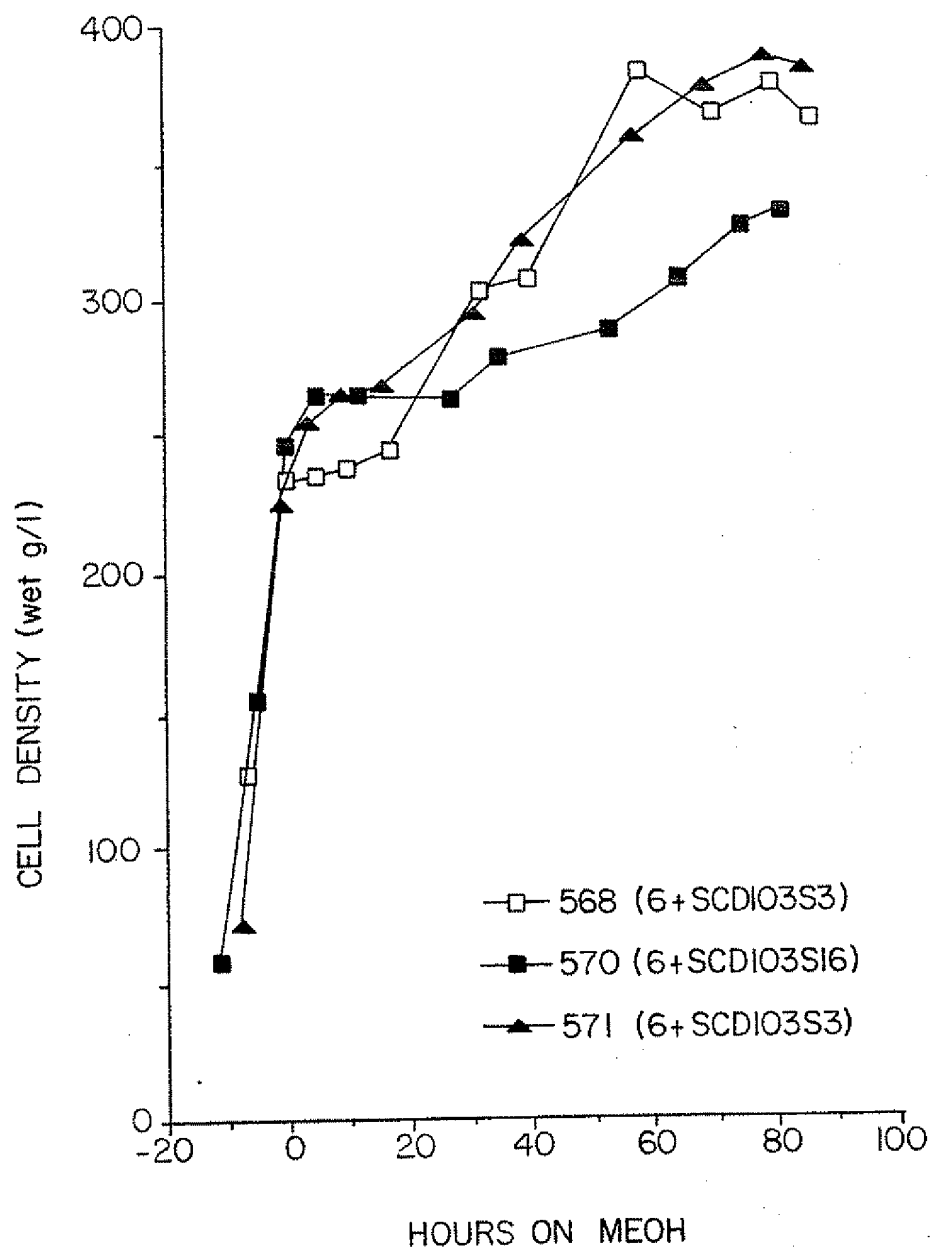


FIG. 6

9 / 15

FIG. 7A

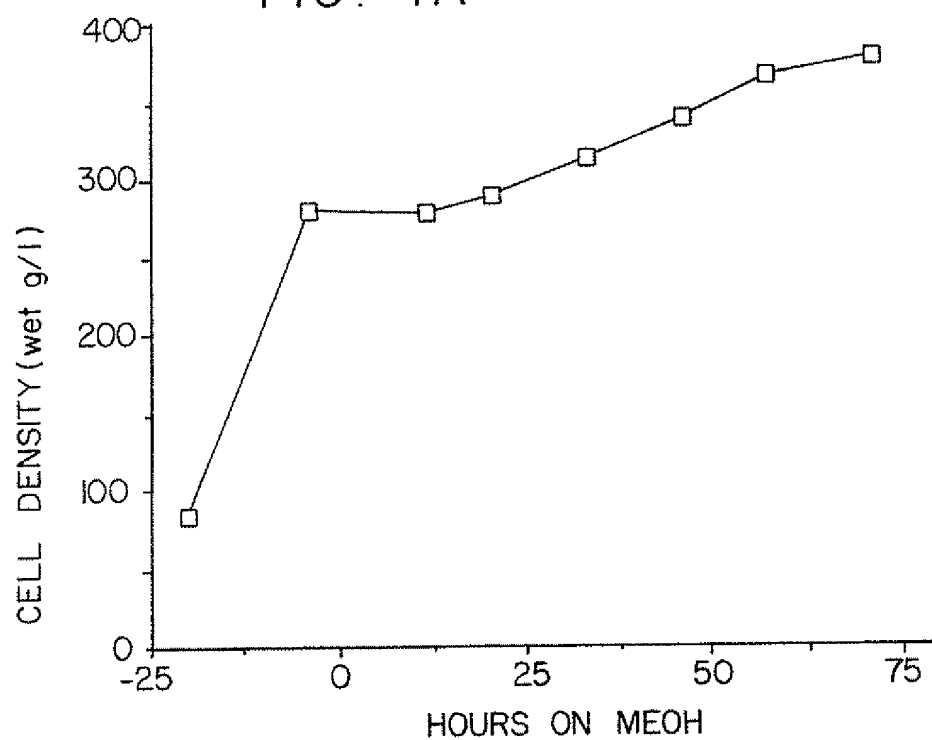
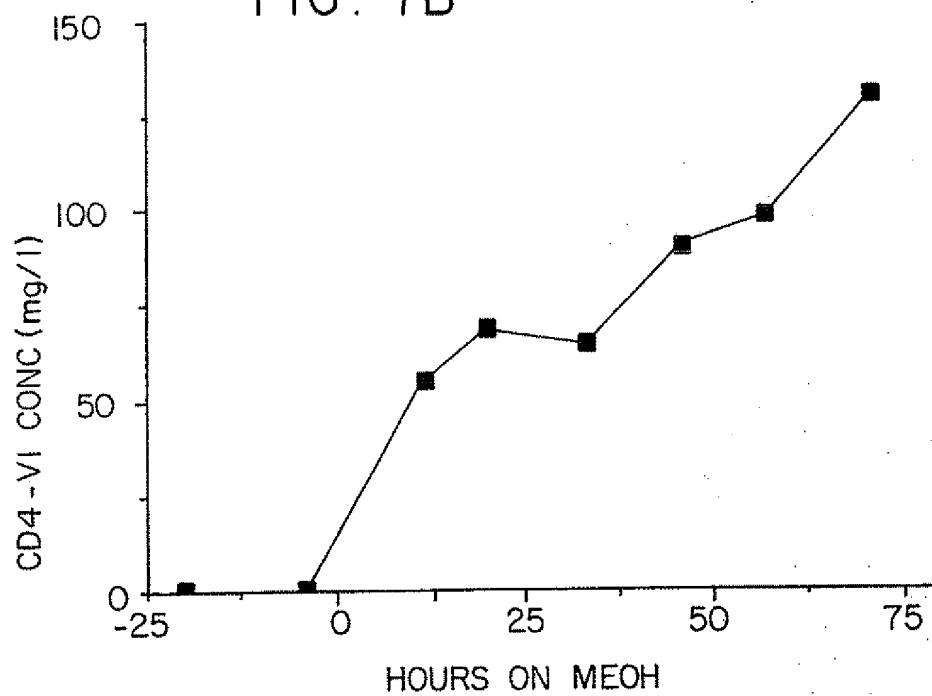


FIG. 7B



SUBSTITUTE SHEET

10 / 15

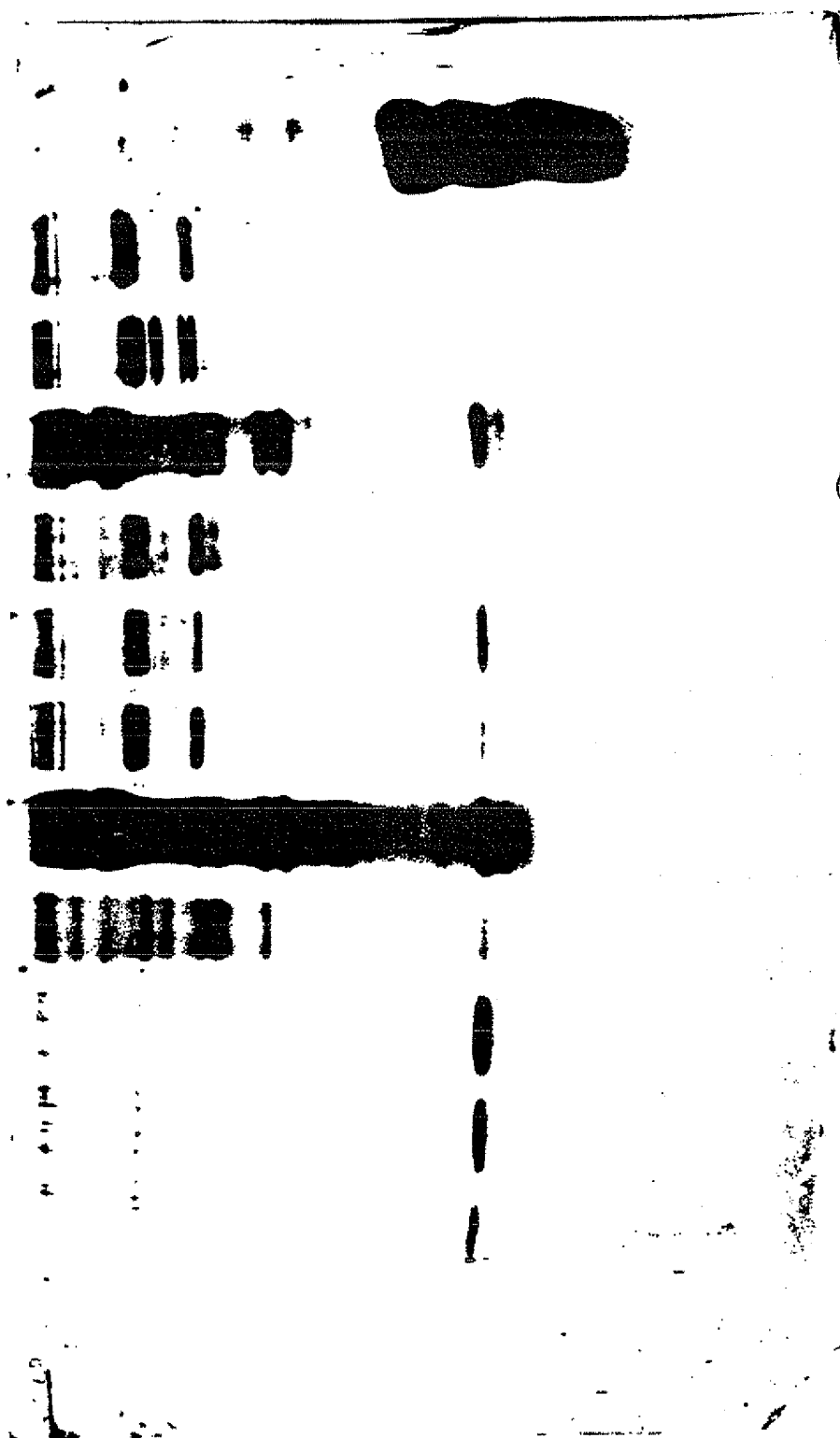


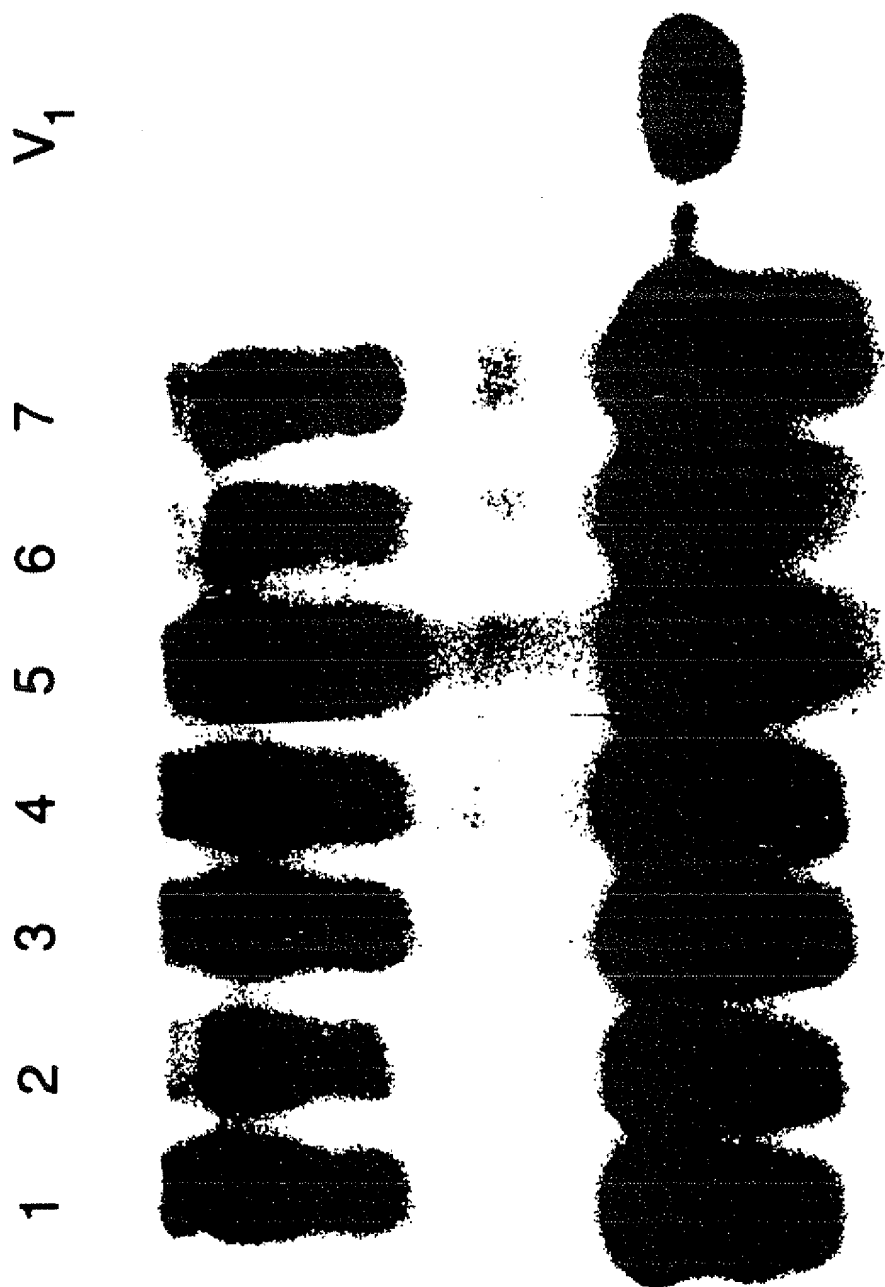
FIG. 8

11 / 15

rCD4-V1	KKVVLGKKGDTVELT
mature CD4	KKVVLGKKGDTVELT

FIG. 9

12 / 15



13 / 15

CAAGCCCAGAGCCCTGCCATTTCTGTGGGCTCAGGTCCCTACTGCTCAGCCCCCTTCCTCC

CTCGGCAAGGCCACA ⁻²⁰ met asn arg gly val pro phe arg his leu leu 108
 ATG AAC CGG GGA GTC CCT TTT AGG CAC TTG CTT

leu val ⁻¹⁰ leu gln leu ala leu leu pro ala ala ^{-1 +1} thr gln gly asn
 CTG GTG CTG CAA CTG GCG CTC CTC CCA GCA GCC ACT CAG GGA AAC

lys val val leu gly lys ⁺¹⁰ lys gly asp thr val glu leu thr cys * 198
 AAA GTG GTG CTG GGC AAA AAA GGG GAT ACA GTG GAA CTG ACC TGT

thr ⁺²⁰ ala ser gln lys lys ser ile gln phe his ⁺³⁰ trp lys asn ser
 ACA GCT TCC CAG AAG AAG AGC ATA CAA TTC CAC TGG AAA AAC TCC

asn gln ile lys ile leu gly ⁺⁴⁰ asn gln gly ser phe leu thr lys
 AAC CAG ATA AAG ATT CTG GGA AAT CAG GGC TCC TTC TTA ACT AAA 288

gly ⁺⁵⁰ pro ser lys leu asn asp arg ala asp ser ⁺⁶⁰ arg arg ser leu
 GGT CCA TCC AAG CTG AAT GAT CGC GCT GAC TCA AGA AGA AGC CTT

trp asp gln gly asn phe pro leu ile ile lys ⁺⁷⁰ asn leu lys ile
 TGG GAC CAA GGA AAC TTC CCC CTG ATC ATC AAG AAT CTT AAG ATA 378

glu ⁺⁸⁰ asp ser asp thr tyr ile cys glu val glu ⁺⁹⁰ asp gln lys glu
 GAA GAC TCA GAT ACT TAC ATC TGT GAA GTG GAG GAC CAG AAG GAG

glu val gln leu leu val phe gly leu thr ala ⁺¹⁰⁰ asn ser asp thr
 GAG GTG CAA TTG CTA GTG TTC GGA TTG ACT GCC AAC TCT GAC ACC 468

his ⁺¹¹⁰ leu leu gln gly gln ser leu thr leu thr ⁺¹²⁰ leu glu ser pro
 CAC CTG CTT CAG GGG CAG AGC CTG ACC CTG ACC TTG GAG AGC CCC

pro gly ser ser pro ser val gln cys arg ser pro arg gly lys
 CCT GGT AGT AGC CCC TCA GTG CAA TGT AGG AGT CCA AGG GGT AAA 558

FIG. 11-1

14 / 15

	+140												+150						
asn	ile	gln	gly	gly	lys	thr	leu	ser	val	ser	gln	leu	glu	leu					
AAC	ATA	CAG	GGG	GGG	AAG	ACC	CTC	TCC	GTG	TCT	CAG	CTG	GAG	CTC					
								+160 *											
gln	asp	ser	gly	thr	trp	thr	cys	thr	val	leu	gln	asn	gln	lys					
CAG	GAT	AGT	GGC	ACC	TGG	ACA	TGC	ACT	GTG	TTG	CAG	AAC	CAG	AAG	648				
+170												+180							
lys	val	glu	phe	lys	ile	asp	ile	val	val	leu	ala	phe	gln	lys					
AAG	GTG	GAG	TTC	AAA	ATA	GAC	ATC	GTG	GTG	CTA	GCT	TTC	CAG	AAG					
								+190											
ala	ser	ser	ile	val	tyr	lys	lys	glu	gly	glu	gln	val	glu	phe					
GCC	TCC	AGC	ATA	GTG	TAT	AAG	AAA	GAG	GGG	GAA	CAG	GTG	GAG	TTC	738				
+200												+210							
ser	phe	pro	leu	ala	phe	thr	val	glu	lys	leu	thr	gly	ser	gly					
TCC	TTC	CCA	CTC	GCC	TTT	ACA	GTT	GAA	AAG	CTG	ACG	GGC	AGT	GGC					
								+220											
glu	leu	trp	trp	gln	ala	glu	arg	ala	ser	ser	ser	lys	ser	trp					
GAG	CTG	TGG	TGG	CAG	GCG	GAG	AGG	GCT	TCC	TCC	TCC	AAG	TCT	TGG	828				
+230												+240							
ile	thr	phe	asp	leu	lys	asn	lys	glu	val	ser	val	lys	arg	val					
ATC	ACC	TTT	GAC	CTG	AAG	AAC	AAG	GAA	GTG	TCT	GTA	AAA	CGG	GTT					
								+250											
thr	gln	asp	pro	lys	leu	gln	met	gly	lys	lys	leu	pro	leu	his					
ACC	CAG	GAC	CCT	AAG	CTC	CAG	ATG	GGC	AAG	AAG	CTC	CCG	CTC	CAC	918				
+260												+270							
leu	thr	leu	pro	gln	ala	leu	pro	gln	tyr	ala	gly	ser	gly	asn					
CTC	ACC	CTG	CCC	CAG	GCC	TTG	CCT	CAG	TAT	GCT	GGC	TCT	GGA	AAC					
								+280											
leu	thr	leu	ala	leu	glu	ala	lys	thr	gly	lys	leu	his	gln	glu					
CTC	ACC	CTG	GCC	CTT	GAA	GCG	AAA	ACA	GGA	AAG	TTG	CAT	CAG	GAA	1008				
+290												+300							
val	asn	leu	val	val	met	arg	ala	thr	gln	leu	gln	lys	asn	leu					
GTG	AAC	CTG	GTG	GTG	ATG	AGA	GCC	ACT	CAG	CTC	CAG	AAA	AAT	TTG					
								+310											
thr	cys	glu	val	trp	gly	pro	thr	ser	pro	lys	leu	met	leu	ser					
ACC	TGT	GAG	GTG	TGG	GGA	CCC	ACC	TCC	CCT	AAG	CTG	ATG	CTG	AGC	1098				

FIG. 11-2

15 / 15

[illegible]

FIG. 11-3

INTERNATIONAL SEARCH REPORT

International Application No PCT/US90/05520

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ²		
According to International Patent Classification (IPC) or to both National Classification and IPC IPC(5): C12P 21/00; C12 N 15/00; C07K 13/00; C07H 15/12 US: 435/69.1, 172.3, 255,320; 530/352; 536/27		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁴		
Classification System	Classification Symbols	
U.S.	435/69.1, 172.3, 255,320; 530/350; 536/27	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁵		
DIALOG BIOTECH FILE (1969-present); CAS FILE (1969-present)		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ¹⁴		
Category [*]	Citation of Document, ¹⁶ with indication, where appropriate, of the relevant passages ¹⁷	Relevant to Claim No. ¹⁸
Y	Bio/Technology, Volume 5, issued May 1987, J.M. Cregg, et al. "High-level expression and efficient assembly of hepatitis B surface antigen in the methylo- trophic yeast, <i>Pichia pastoris</i> ", pages 479-485, see entire article.	1-40
Y X	Proceedings of the National Academy of Sciences, Volume 85, issued April 1988. E.A. Berger, et al. "A soluble recombinant polypeptide comprising the amino-terminal half of the extracellular region of the CD4 molecule contains an active binding site for human immunodeficiency virus", pages 2357-2361, see entire article.	1-38, 40 39
Y	US, A, 4,546,082 (Kurjan, et al.) 08 October 1985, see entire document.	1-40
Y	US, A, 4,808,534 (Schick, et al.) 28 February 1989, see entire document.	32-38
Y	US, A, 3,929, 578 (Urakami) 30 December 1975, see entire document.	32-38
<p>[*] Special categories of cited documents: ¹⁵</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such docu- ments, such combination being obvious to a person skilled in the art.</p> <p>"A" document member of the same patent family</p>		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search ²	Date of Mailing of this International Search Report ²	
29 OCTOBER 1990	25 JAN 1991	
International Searching Authority ¹	Signature of Authorized Officer ²⁰	
ISA/US	BETH A. BURROUS	